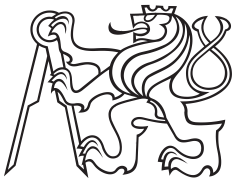
Bakalářská práce



České
vysoké
učení technické
v Praze

F3

Fakulta elektrotechnická
Katedra kybernetiky

Směrové statistiky v predikci kvaziperiodických časových řad

Leonard Mentzl

Vedoucí: prof. Ing. Mirko Navara, DrSc.
Obor: Informatika a počítačové vědy
Studijní program: Otevřená informatika
Květen 2020

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Mentzl** Jméno: **Leonard** Osobní číslo: **474391**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra kybernetiky**

Studijní program: **Otevřená informatika**

Studijní obor: **Informatika a počítačové vědy**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Směrové statistiky v predikci kvaziperiodických časových řad

Název bakalářské práce anglicky:

Directional Statistics for Prediction of Quasiperiodic Time Series

Pokyny pro vypracování:

Seznamte se s metodami odhadu parametrů pravděpodobnostních modelů (metoda maximální věrohodnosti, metoda momentů atd.) pro data, která jsou číselná, ale s jinou než obvyklou aritmetikou, např. úhly. Navrhněte alespoň jeden vhodný model a postup stanovení jeho parametrů a ověřte jeho použitelnost pro popis a predikci časových řad s periodickou složkou, např. na datech z článku [4].

Seznam doporučené literatury:

[1] Schlesinger, M.I., Hlaváč, V.: Deset přednášek z teorie statistického a strukturního rozpoznávání. ČVUT, Praha, 1999.
[2] Papoulis, A., Pillai, S.U.: Probability, Random Variables, and Stochastic Processes. 4th ed., McGraw-Hill, Boston, USA, 2002.
[3] Wasserman, L.: All of Statistics. A Concise Course in Statistical Inference. Springer, 2004.
[4] Krajník, T. et al.: Warped Hypertime Representations for Long-term Autonomy of Mobile Robots. IEEE Robotics and Automation Letters. accepted June, 2019.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

prof. Ing. Mirko Navara, DrSc., Strojové učení FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **10.12.2019** Termín odevzdání bakalářské práce: **22.05.2020**

Platnost zadání bakalářské práce: **30.09.2021**

prof. Ing. Mirko Navara, DrSc.	doc. Ing. Tomáš Svoboda, Ph.D.	prof. Mgr. Petr Páta, Ph.D.
podpis vedoucí(ho) práce	podpis vedoucí(ho) ústavu/katedry	podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

. Datum převzetí zadání

Podpis studenta

© ČVUT v Praze, Design: ČVUT v Praze, VIC

Poděkování

Děkuji ČVUT, že mi je tak dobrou *alma mater*.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Mirošovicích, 15. května 2020

Abstrakt

V této práci navrhujeme metody predikce kvaziperiodických časových řad, které jsou založené na směrové statistice. Zabýváme se metodou momentů, EM algoritmem, Adaboostem$_{\mathbb{R}}$ a neuronovými sítěmi. Upřednostňujeme korektnost a vyhýbáme se používání heuristických metod. Hledáme zcela nové modely a metody, takže jejich vlastnosti ukazujeme na jakýchkoli úlohách, včetně umělých, které ukazují, co by se mohlo stát, bez ohledu na to, zda to v konkrétní aplikaci opravdu hraje roli. Navržené metody testujeme na reálných datech a porovnáváme je s metodami: odhad průměrem, histogram, FreMEn a HyT-EM. Výsledky ukazují, že pro některá data jsou některé navržené metody statisticky významně lepší než metody FreMEn a HyT-EM.

Klíčová slova: směrová statistika, chronorobotika, metoda momentů, EM algoritmus, logistická regrese, Adaboost, neuronové sítě

Vedoucí: prof. Ing. Mirko Navara, DrSc.

Abstract

In this thesis, we propose new methods for prediction of quasiperiodic time series which are based on directional statistics. We study the moment matching, EM algorithm, Adaboost$_{\mathbb{R}}$ and neural networks. We emphasize correctness of the methods and avoid usage of heuristics. We search for completely new methods so we are showing their features on any tasks, including the artificial ones, which show what might happen regardless of the fact whether it actually plays a role in the specific application. We test the proposed methods on real data and compare them with the following methods: estimation by mean, histogram, FreMEn and HyT-EM. The results show that some of the methods are statistically significantly better for some data than methods FreMEn and HyT-EM.

Keywords: directional statistics, chronorobotics, moment matching, EM algorithm, Adaboost, logistic regression, neural networks

Title translation: Directional Statistics for Prediction of Quasiperiodic Time Series

Obsah

1 Úvod 1

1.1 Motivace 1

1.2 Jevy v predikci kvaziperiodických časových řad 2

1.3 Použité značení 3

2 Zkoumané úlohy 5

2.1 Úloha s dveřmi: Odhad jedné binární veličiny 5

2.2 Úloha s vytížeností: Odhad jedné diskrétní veličiny 6

2.3 Úloha s chodci: Odhad pravděpodobnostního rozdělení v závislosti na čase 6

2.4 Srovnávací kritérium 8

3 Přehled již zavedených metod 9

3.1 Odhad průměrem 9

3.2 Histogram 9

3.3 FreMEn [8] 9

3.4 HyT-EM [9] 10

3.5 Gravitační shlukování 10

4 Kruhové momenty 13

5 Nástin řešení zkoumaných úloh 17

5.1 Úloha s dveřmi 17

5.2 Úloha s vytížeností 17

5.3 Úloha s chodci 18

6 Použitá rozdělení na kružnici 19

6.1 Von Misesovo rozdělení 19

6.2 Kruhové Cauchyovo rozdělení . . 20

6.3 Odhad parametrů 21

7 Odhad rozdělení na kružnici pomocí metody momentů (MM-vM a MM-kC) 25

7.1 Metoda momentů a směs von Misesových rozdělení (MM-vM) . . . 26

7.2 Metoda momentů a směs kruhových Cauchyových rozdělení (MM-kC) 27

8 Odhad rozdělení na kružnici pomocí EM algoritmu (EM-vM a EM-kC) 29

8.1 EM algoritmus pro směs von Misesových rozdělení (EM-vM) ... 29

8.2 EM algoritmus pro směs kruhových Cauchyových rozdělení (EM-kC) .. 31

9 Logistická regrese pro von Misesovo rozdělení 33

9.1 Logistická regrese pro von Misesovo-Fisherovo rozdělení 34

10 Adaboost$_{\mathbb{R}}$ [13] 37

11 Neuronové sítě (NS) 39

11.1 Směrová statistika a fuzzy logika v neuronové síti 41

11.2 Vlastnosti neuronové sítě 44

12 Aplikace metod na úlohu s dveřmi 47

12.1 Metoda momentů (MM-vM a MM-kC) 47

12.2 EM algoritmus (EM-vM a EM-kC) 47

12.3 Adaboost$_{\mathbb{R}}$ 48

12.4 Neuronová síť (NS) 48

12.5 Srovnání výsledků 48

13 Aplikace metod na úlohu s vytížeností 55

13.1 Metoda momentů (MM-vM a MM-kC) 55

13.2 EM algoritmus (EM-vM a EM-kC) 55

13.3 Adaboost$_{\mathbb{R}}$ 56

13.4 Neuronová síť (NS) 56

13.5 Srovnání výsledků 56

14 Aplikace metod na úlohu s chodci 61

14.1 Metoda momentů (MM-vM a MM-kC) 61

14.2 EM algoritmus (EM-vM a EM-kC) 61

14.3 Adaboost$_{\mathbb{R}}$ 62

14.4 Neuronová síť (NS) 62

14.5 Srovnání výsledků 62

15 Závěr 67

15.1 Metoda momentů 67

15.2 EM algoritmus 68

15.3 Adaboost$_{\mathbb{R}}$ 68

15.4 Neuronové sítě 68

Literatura **69**

A Důkaz některých vlastností kruhového Cauchyova rozdělení **71**

Obrázky

2.1 Trénovací data k úloze s dveřmi . 6

2.2 Trénovací data k úloze s vytížeností 7

2.3 Trénovací data k úloze s chodci. Pro přehlednost byla zobrazena pouze padesátina měření. 7

4.1 Kružnice přetransformovaná mocninami 15

6.1 Hustota von Misesova rozdělení pro $\mu = 0$ a $\kappa \in \{0; 1; 2; 4\}$ v polárních souřadnicích. Argumentem je úhel a hustota je znázorněna vzdáleností od počátku. 20

6.2 Hustota kruhového Cauchyova rozdělení pro $\mu \in \{0; 0.3; 0.5; 0.7\}$ v polárních souřadnicích. Argumentem je úhel a hustota je znázorněna vzdáleností od počátku. 21

6.3 Odhad parametru μ z 10, 20, 30, 50 a 100 bodů. 23

6.4 Odhad parametru μ z ilustračních dat 23

7.1 Odhad pěti shluků pomocí metody momentů jako směs von Misesových rozdělení. Data jsou pouze ilustrační, byla vytvořena z dat [9]. 26

7.2 Odhad pěti shluků pomocí metody momentů jako směs dvou (vlevo) a třech (vpravo) kruhových Cauchyových rozdělení. Na obrázku vpravo vidíme, že model sice danou soustavu řeší, ale neodpovídá datům. Data jsou pouze ilustrační, byla vytvořena z dat [9]. 27

8.1 Odhad pěti shluků pomocí EM algoritmu jako směs von Misesových rozdělení (vlevo) a jako směs kruhových Cauchyových rozdělení (vpravo). Data jsou pouze ilustrační, byla vytvořena z dat [9]. 30

9.1 Příklad logistické regrese. Červené kroužky značí měření první třídy, modré křížky značí měření druhé třídy. Vrstevnice vyjadřují odhad aposteriorní pravděpodobnosti první třídy. Data jsou pouze ilustrační. . 34

12.1 Úloha s dveřmi: predikce použitých metod pro trénovací sadu 49

12.2 Úloha s dveřmi: predikce použitých metod pro testovací sadu 6 50

12.3 Úloha s dveřmi: střední kvadratická chyba pro metody s parametry 51

12.4 Úloha s dveřmi: Šipka z metody A do metody B znamená, že metoda A je lepší než metoda B podle kritéria uvedeného v kapitole 2.4. . 52

13.1 Úloha s vytížeností: predikce použitých metod pro trénovací sadu. Predikce metod EM-vM a MM-kc téměř splývají. 57

13.2 Úloha s vytížeností: predikce použitých metod pro testovací sadu 4. Predikce metod EM-vM a MM-kc téměř splývají. 58

13.3 Úloha s vytížeností: střední kvadratická chyba pro metody s parametry 59

13.4 Úloha s vytížeností: Šipka z metody A do metody B znamená, že metoda A je lepší než metoda B podle kritéria uvedeného v kapitole 2.4. 60

14.1 Úloha s chodci: střední kvadratická chyba pro metody s parametry 63

14.2 Úloha s chodci: Šipka z metody A do metody B znamená, že metoda A je lepší než metoda B podle kritéria uvedeného v kapitole 2.4. . 64

Tabulky

12.1 Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro jednu periodicitu 53

12.2 Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro dvě periodicity . . 53

12.3 Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro tři periodicity . . . 54

12.4 Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro čtyři periodicity . 54

14.1 EM-vM: seznam nalezených shluků 65

14.2 EM-kC: seznam nalezených shluků 66

ctuthesis t1606152353

Kapitola 1

Úvod

1.1 Motivace

Máme následující případ: Máme herní server, jehož vytíženost se mění v závislosti na denní době a dnu v týdnu. Abychom dosáhli plného využití serveru, můžeme mimo špičky používat některé prostředky k jiným účelům. K tomu by bylo dobré umět předem předpovědět, jak bude server vytížen. Podobně můžeme modelovat vytíženost jiných veřejných zařízení, abychom věděli například, kolik prostředků máme nakoupit předem, kdy máme provádět údržbu, a podobně. Toto má význam pro internetové obchody, zejména s potravinami. Můžeme také zdražit služby ve špičkách, ale potom se už nejedná o rozpoznávání, ale o hru, protože zdražení služeb zapříčiní změnu vytíženosti.

Máme ještě jednu podobnou úlohu. Hustota aut na vozovce se mění v závislosti na denní periodě. Naším úkolem je navrhnout metodu, která dokáže předpovědět hustotu aut v daném čase a případně ještě předpovědět vektor rychlosti. S touto znalostí je možné naplánovat cestu městem takovou, aby jízda byla co nejrychlejší s ohledem na vytíženost jednotlivých komunikací.

„Pokroky v autonomní robotice postupně umožňují nasazení robotů v lidském prostředí [6]. Lidská aktivita mění prostředí a roboti, kteří sdílejí prostředí s lidmi, se musí s těmito neustálými změnami vyrovnat. Spousta autorů ukázala, že modely prostředí, které se přizpůsobují změnám, zlepšují celkovou schopnost mobilních robotů pracovat dlouhodobě [2, 4, 6, 8, 10, 16]. Jelikož dlouhodobé nasazení robota způsobuje, že změny prostředí jsou po-

zorovány častěji, mobilní roboti získávají možnost naučit se nejen strukturu prostředí, ale také, jak se mění v závislosti na čase.“ [9]

Ukazuje se, že modely, které modelují čas periodicky, úspěšně modelují změny v lidském prostředí [8, 9]. Je zřejmé, že zde hraje roli denní a týdenní perioda, v menší míře i další periody. Měření se během periody výrazně mění, ale každá perioda je skoro stejná. Proto se v této práci hlouběji zaměřujeme na periodické časové modely se známou periodou za účelem objevení dosud neprozkoumaných možností.

Směrová statistika je obor statistiky, který se zabývá pravděpodobnostními rozděleními na prostorech jednotkových vektorů, které reprezentují směr, a unitárních zobrazení. Směrovou statistiku také můžeme využít při zpracování periodických dat, pokud si jednu periodu zobrazíme na jednotkovou kružnici. Tím se konec periody zobrazí poblíž začátku periody. Něčeho podobného využívá metoda HyT-EM [9], ale používá nekorektní statistický model, protože modeluje pravděpodobnostní rozdělení na křivce pomocí směsi vícerozměrných Gaussových rozdělení. Otevírá se tady dosud neprozkoumaná oblast, a sice predikce kvaziperiodických časových řad pomocí směrové statistiky.

1.2 Jevy v predikci kvaziperiodických časových řad

Při predikci kvaziperiodických časových řad může dojít k těmto typickým jevům:

Překmitnutí je predikce hodnoty, která vybočuje z požadovaných mezí, případně z mezí, ve kterých se měřené hodnoty běžně vyskytují. V predikci kvaziperiodických časových řad se snažíme, aby k překmitávání nedocházelo, případně k němu docházelo co nejméně. Tento jev můžeme odstranit tak, že předpovídané hodnoty vhodně ořízneme. Nicméně se jedná pouze o heuristiku a může nastat problém ve chvíli, kdy meze nejsou předem dané a musíme je odhadovat.

Náhlá změna je jev v trénovacích datech, který odráží nějakou změnu v reálném světě, při které došlo ke změně rozvrhu událostí. Tou může být například změna otevírací doby v obchodě.

Výjimečná událost je jev v trénovacích datech, který odráží nějakou událost v reálném světě, která nezapadá do běžného rozvrhu událostí. Tou může být

například oprava vodovodního potrubí. Výjimečnou událost můžeme rovněž vnímat jako dvě náhlé změny jdoucí bezprostředně za sebou. Nejedná se o vychýlené měření, protože vychýlené měření vypovídá o něčem, co se reálně nestalo, zatímco výjimečná událost vypovídá o něčem, co se skutečně stalo.

Zapomínání je odolnost prediktoru vůči náhlým změnám a výjimečným událostem. Prediktor měl patřičně zohledňovat data, která byla naměřena před nějakou náhlou změnou nebo během výjimečné události. Pokud dojde ke změně rozvrhu událostí, prediktor by se měl co nejrychleji přeladit na nový rozvrh. Také může jít o postupnou změnu, např. v chování lidí, jejich denním režimu apod. Pokud dojde k nějaké výjimečné události, pak by ji prediktor neměl začlenit do běžného rozvrhu. Odolnost vůči náhlým změnám můžeme řešit tak, že budeme zahazovat stará data, nebo jim budeme přikládat menší důležitost. Problém je, že se jedná pouze o heuristiku a nevíme, jak stará data máme zahazovat či jak velkou důležitost máme přikládat starým datům. Existují i jiné modely zapomínání, jako třeba exponenciální zapomínání, ale všechny se spoléhají na nějaký parametr, který nakonec musíme určit heuristicky. Navíc se tím snižuje robustnost, odolnost vůči výjimečným událostem a vylučuje se tím možnost vytvoření delšího časového modelu.

1.3 Použité značení

Symbol ϑ označuje bod na jednotkové kružnici v komplexní rovině, zatímco symbol θ označuje úhel, který je v intervalu $(-\pi;\pi\rangle$. Mezi těmito symboly existuje jednoduchý převodní vztah:

$$\vartheta = \mathrm{e}^{\theta i} \tag{1.1}$$

Podobné značení bylo zvoleno proto, že tyto dvě hodnoty mají velice podobný význam, ale nejde jedna zaměňovat za druhou. Pro větší přehlednost nikdy nebudou v jednom výrazu použity obě, vždy jen jedna, podle toho, která se pro daný výpočet zrovna lépe hodí.

Symbol Γ pod integrálem značí jednotkovou kružnici. Pokud se uvnitř integrálu vyskytuje symbol θ, myslí se tím interval $(-\pi;\pi\rangle$, pokud se uvnitř integrálu vyskytuje symbol ϑ, myslí se tím jednotková kružnice v komplexní rovině. Symbol $\mathrm{d}p$ značí, že integrujeme podle pravděpodobnostní míry daného rozdělení, a tedy se výrok týká spojitého, diskrétního i smíšeného rozdělení. A protože integrujeme přes pravděpodobnostní míru, tak z definice pravděpodobnosti plyne:

$$\int_\Gamma 1 \,\mathrm{d}p = 1 \tag{1.2}$$

Kapitola 2

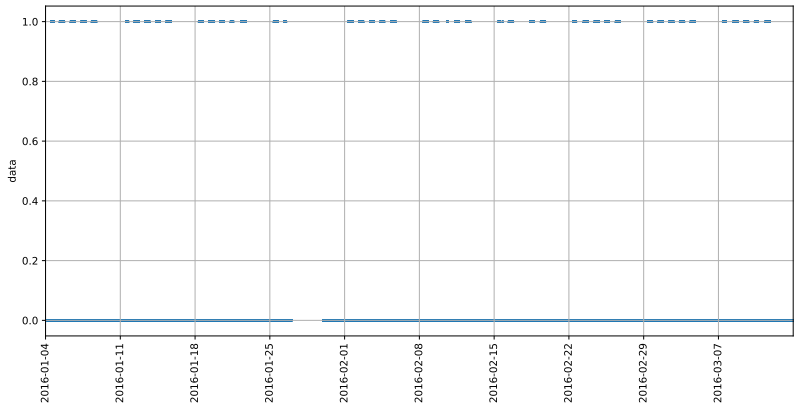
Zkoumané úlohy

Byly zkoumány tři úlohy na reálných datech. Ve všech případech smíme předpovídat libovolné reálné hodnoty, i když někdy nemohou logicky nastat. Nicméně se tím může snížit součet čtverců odchylek. Pokud není uvedeno jinak, navržené metody testujeme na datech z [9]. Úloha s lokalizací byla vynechána kvůli implementační chybě, kterou se nepodařilo odstranit.

2.1 Úloha s dveřmi: Odhad jedné binární veličiny

Po určitou dobu byly sledovány dveře do kanceláře a každou minutu bylo zapsáno, zda-li jsou dveře otevřené či ne, tedy 1 nebo 0. Úkolem je pro daný čas odhadnout, zda-li budou dveře otevřené či ne. Snažíme se minimalizovat součet čtverců odchylek v časech, kdy byla pořízena testovací měření.

Dveře byly pozorovány RGB-D kamerou po dobu 10 týdnů pro získání trénovací sady a dalších 9 týdnů pro získání devíti testovacích sad, každá je dlouhá jeden týden. V každém testovacím týdnu je prvních pět dní všedních a poslední dva jsou víkendové. Protože zpracovávání dat RGB-D kamerou bylo spíše jednoduché, data obsahují šum, protože lidé procházející dveřmi způsobovali, že dveře byly nesprávně detekovány jako zavřené.



Obrázek 2.1: Trénovací data k úloze s dveřmi

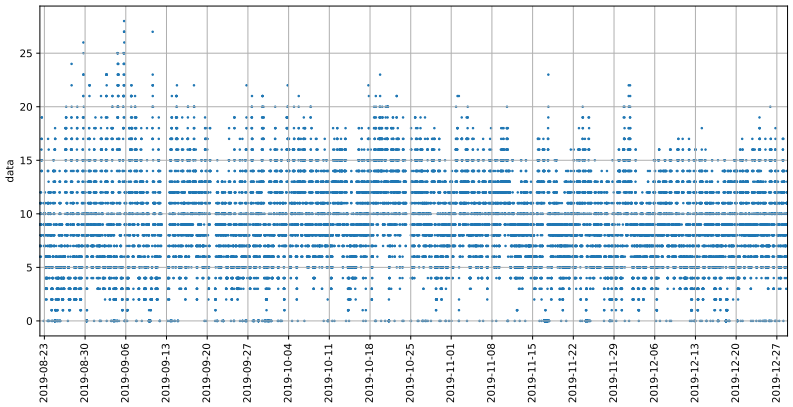
2.2 Úloha s vytížeností: Odhad jedné diskrétní veličiny

Známe historii počtu připojených hráčů na nějakém herním serveru. Úkolem je předpovědět počet připojených hráčů v budoucnosti. Snažíme se minimalizovat součet čtverců odchylek v časech, kdy byla pořízena testovací měření.

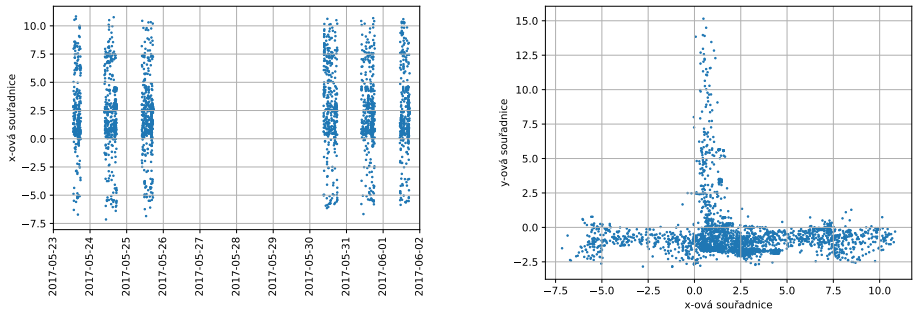
Data byla získána ze záznamů serveru Tunnelers' Abyss se souhlasem jeho provozovatele. Získaná data jsou anonymní, obsahují pouze počty hráčů a časy. Používáme data z období od 22. 8. 2019 do 1. 2. 2020, pro každých 10 minut je jeden záznam. Posledních 50000 záznamů bylo vyčleněno pro testování, každou testovací sadu tvoří 10000 po sobě jdoucích záznamů.

2.3 Úloha s chodci: Odhad pravděpodobnostního rozdělení v závislosti na čase

Řešíme následující úlohu: Po určitou dobu jsme sledovali chodce na chodbě. Potřebujeme pro dané časoprostorové okno určit, kolik chodců se v něm ocitne. Snažíme se minimalizovat součet čtverců odchylek v časech, kdy byla pořízena testovací měření.



Obrázek 2.2: Trénovací data k úloze s vytížeností



Obrázek 2.3: Trénovací data k úloze s chodci. Pro přehlednost byla zobrazena pouze padesátina měření.

Použitá data byla pořízena na univerzitě v Lincolnu, v Isaan Newton Building. Sběr dat byl proveden pohyblivým robotem s laserem Velodyne 3d, který byl umístěn na téčkovém křížení dvou chodeb. Na detekci chodců byla použita metoda [17]. Data nebyla snímána nepřetržitě, protože robot se musel dobíjet nebo řešit jiné úkoly.

Tato úloha má význam mimo jiné i v situaci, kdy je mezi lidmi rozšířena nějaká epidemie a je nutné zabránit větší koncentraci lidí. Z časového modelu je potom možné navrhnout taková opatření, která zamezí shlukování lidí a tím i sníží riziko přenosu epidemie.

2.4 Srovnávací kritérium

Srovnání metod probíhá ve dvou kolech. V prvním kole porovnáváme stejné metody s různými parametry a vybíráme ten nejlepší parametr. Ve druhém kole porovnáváme metody s nejlepšími parametry mezi sebou. Tento postup volíme proto, aby výsledné grafy byly přehledné a neobsahovaly příliš mnoho údajů.

Pokud daná metoda má parametr, který vybíráme z předem stanovené množiny možností (např. počet shluků), natrénujeme metodu pro všechny možnosti. Tento parametr potom uvádíme jako číslo za jménem metody. Pro každou z těchto možností vypočítáme střední kvadratickou chybu na každé testovací množině a tyto chyby sečteme. Vybereme tu volbu parametru, pro který byl tento součet nejmenší, a metoda s tímto parametrem postupuje do druhého kola. Pokud to vyjde nerozhodně, jako sekundární kritérium použijeme jednoduchost modelu.

Ve druhém kole metody mezi sebou porovnáváme tak, že mezi sebou odečteme střední kvadratické chyby jedné a druhé metody odpovídající stejným testovacím množinám. Pomocí Studentova t-testu otestujeme na hladině významnosti $\alpha = 0.95$, zda mají rozdíly střední hodnotu nulovou. Pokud test zamítne hypotézu, že rozdíly mají střední hodnotu nezápornou, řekneme, že druhá metoda je lepší. Pokud test zamítne hypotézu, že rozdíly mají střední hodnotu nekladnou, řekneme, že první metoda je lepší. Pokud test nezamítne žádnou z předchozích dvou hypotéz, řekneme, že tyto dvě metody jsou nesrovnatelné.

Metody srovnáváme tímto způsobem, protože stejný způsob byl zvolen v článku [9]. Bylo by dobré při srovnání uvažovat i složitosti modelů, čas nutný k vytvoření modelu a čas nutný k výpočtu předpovědi. To už se ale jedná o multikriteriální optimalizaci a v každé aplikaci se zmíněná kritéria uplatňují jiným způsobem. Jelikož se v této práci nezabýváme žádnou konkrétní aplikací, neexistuje logický způsob, jak zbývající kritéria využít. Tato kritéria hodnotíme pouze slovně a srovnání ponecháváme pro případného uživatele těchto metod.

Kapitola 3

Přehled již zavedených metod

3.1 Odhad průměrem

Tato metoda je jednoduchá. Prostě vezmeme všechna měření a vypočítáme aritmetický průměr. Čas nikterak nezohledňujeme. Pokud předpovídáme více údajů, tak jednotlivé údaje předpovídáme nezávisle na sobě.

3.2 Histogram

Předem si stanovíme periodu a tu si rozdělíme rovnoměrně na daný počet dílů. Pro každý díl vypočítáme aritmetický průměr, a ten použijeme jako odhad pro časy, které spadají do daného dílu v periodě. Pokud předpovídáme více údajů, tak jednotlivé údaje předpovídáme nezávisle na sobě.

3.3 FreMEn [8]

Na datech provedeme diskrétní Fourierovu transformaci. Z ní vybereme předem stanovený počet hodnot s největší absolutní hodnotou a zbytek nasta-

víme na nulu. Inverzní Fourierovou transformací dostaneme předpověď. Pokud máme zadané meze, tak výsledek ořízneme, aby se pohyboval v požadovaných mezích. Pokud předpovídáme více údajů, tak jednotlivé údaje předpovídáme nezávisle na sobě.

3.4 HyT-EM [9]

Předem si zvolíme počet shluků. Počet period N na začátku nastavíme na 0, tedy v prvním kroku čas neuvažujeme. Každé měření v čase t a v místě x přetransformujeme na:

$$\check{x} = \begin{bmatrix} \cos(2\pi t/T_1) \\ \sin(2\pi t/T_1) \\ \vdots \\ \cos(2\pi t/T_N) \\ \sin(2\pi t/T_N) \\ x \end{bmatrix} \tag{3.1}$$

kde T_1 až T_N jsou zatím nalezené periody. Začínáme s prázdnou množinou period, tedy v prvním kroku $\check{x} = x$. Na těchto přetransformovaných datech provedeme EM algoritmus pro vícerozměrné Gaussovo rozdělení. Zde může být problém, protože poslední souřadnice je prostorová (v jednotkách délky), zatímco ostatní jsou bezrozměrné a z intervalu $\langle -1; 1 \rangle$. Hrozí chyby, pokud se správně nevypořádáme s otázkou volby jednotek délky. Práce [9] tento problém neřeší.

Spočítáme součet čtverců odchylek a podíváme se, jestli je lepší než chyba předchozího modelu. Pokud ne, tak vrátíme model s nejmenší chybou. V opačném případě najdeme nejvýraznější periodicitu chyby pomocí metody FreMEn [8], přidáme ji k nalezeným periodám a opakujeme postup.

Klasifikace probíhá podle toho, jakou úlohu zrovna řešíme. Bližší informace jsou uvedeny v práci [9].

3.5 Gravitační shlukování

Práce [3] zkoumá algoritmy gravitačního shlukování v obecných metrických prostorech. V této práci se také zabýváme shlukováním, ale ne na obecných

metrických prostorech, ale pouze na jednotkové kružnici. Zatímco práce [3] používá gravitační algoritmy, které jsou spíše heuristické, v této práci se snažíme aplikovat statistické metody. Výše zmíněným úlohám se však práce [3] nevěnuje.

Kapitola 4

Kruhové momenty

Pokud uvažujeme pravděpodobnostní rozdělení na kružnici, můžeme počítat momenty více způsoby. Nabízí se ta možnost, že se kružnice rozvine na polouzavřený interval a na něm se počítají momenty klasickým způsobem. Tento způsob není vhodný, protože momenty závisí na tom, ve kterém místě se kružnice rozdělí. Tuto nevýhodu ilustrujeme na několika příkladech. V každém příkladu budeme uvažovat, že jednotlivé body na kružnici odpovídají světovým stranám. Zde budeme pro názornost měřit úhly ve stupních, ve zbytku práce budeme měřit úhly v radiánech.

Uvažujme rovnoměrné rozdělení na kružnici. Pokud se rozhodneme měřit azimut od $-180°$ do $180°$, střední hodnota vyjde na severu, tj. $0°$. Pokud se rozhodneme měřit azimut od $0°$ do $360°$, střední hodnota vyjde na jihu, tj. $180°$. Přitom ani jedna z nabízených hodnot se nezdá být smysluplná.

Nyní uvažujme diskrétní rozdělení na kružnici, které bodu $-90°$ přiřazuje pravděpodobnost $\frac{1}{2}$ a bodu $90°$ pravděpodobnost $\frac{1}{2}$. Snažíme se tedy najít střední hodnotu mezi východním a západním směrem. V závislosti na tom, kde kružnici roztrhneme, vyjde střední hodnota buď $0°$ (tj. sever), nebo $180°$ (tj. jih). Hledat průměr mezi východem a západem může být smysluplná úloha, ale některé postupy dávají zcela zcestné výsledky.

Jako poslední příklad si vezměme rozdělení, které je koncentrované kolem jižního směru. Pokud se rozhodneme měřit azimut od $-180°$ do $180°$, rozdělení se roztrhne na dva shluky, jeden bude poblíž $-180°$ a druhý bude poblíž $180°$. Střední hodnota tedy vyjde někde na severu. Tomu můžeme zamezit tím, že bod předělu zvolíme jinde. Jenomže co máme dělat v případě, kdy neznáme

předem parametry rozdělení? Ať už bod předělu dáme kamkoliv, může nám hrozit, že se zrovna trefíme někam špatně.

Z předchozích příkladů je patrné, že největším kamenem úrazu je bod předělu, takže výsledek není invariantní vůči rotaci, tj. stejný směr má vyjít, bez ohledu na to, jak úlohu otočíme. Z prvního příkladu navíc plyne, že někdy ani není možné bod předělu zvolit tak, aby momenty vycházely smysluplně, a to i když rozdělení předem tušíme. Proto je lepší se úplně vyhnout rozvíjení kružnice do intervalu. Další metody vycházejí z toho, že rozdělení na kružnici budeme považovat za dvourozměrné rozdělení, jehož body se nacházejí na jednotkové kružnici.

Jednotkovou kružnici můžeme reprezentovat v množině $\mathbb{R}^2$ nebo v množině $\mathbb{C}$. V množině $\mathbb{R}^2$ můžeme počítat momenty běžným způsobem. Tomu budeme říkat plošné momenty. V množině $\mathbb{C}$ můžeme počítat momenty stejně jako v množině $\mathbb{R}$, pouze s tím rozdílem, že místo reálných čísel budeme používat komplexní čísla, a tedy výsledkem budou také komplexní čísla. Tomu budeme říkat kruhové momenty. Všimněme si, že sčítání funguje stejně v $\mathbb{R}^2$ a v $\mathbb{C}$. Jelikož k výpočtu prvního momentu není potřeba umocňování, můžeme tvrdit, že pro výpočet střední hodnoty nezáleží na tom, zda reprezentujeme kružnici v $\mathbb{R}^2$, nebo v $\mathbb{C}$. Všimněme si, že pokud počítáme střední hodnotu plošně nebo kruhově, odstraníme tím vady předchozího přístupu.

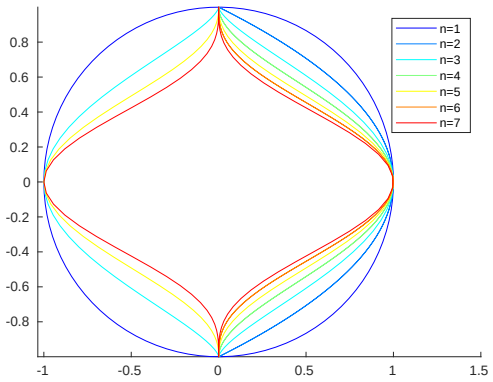
Pokud plošné momenty nebo kruhové momenty aplikujeme na výše uvedené příklady, dostaneme lepší výsledky. V prvních dvou příkladech střední hodnota vyjde nulová. To znamená, že žádný střední směr nemá smysl hledat. Ve třetím příkladu nám střední hodnota vyjde poblíž jihu. Můžeme obecně říci, že pokud je rozdělení na kružnici koncentrované kolem právě jednoho místa, střední hodnota vyjde smysluplně.

Za povšimnutí stojí fakt, že střední hodnota má nyní dva rozměry. Směr střední hodnoty udává, na kterou stranu je rozdělení vychýlené, zatímco absolutní hodnota udává významnost této výchylky. Může se i stát, že je střední hodnota nulová, což znamená, že rozdělení není vychýlené žádným směrem, a tedy nemá smysl mluvit o středním úhlu.

Co se týče vyšších momentů, plošné a kruhové momenty se různí, ale lze převádět jedny na druhé a naopak. Například pro druhý x-ový moment platí:

$$\int_\Gamma \cos^2\theta \,\mathrm{d}p = \int_\Gamma \left(\frac{1}{2} + \frac{1}{2}\cos 2\theta\right) \mathrm{d}p = \frac{1}{2} + \frac{1}{2}\int_\Gamma \cos 2\theta \,\mathrm{d}p \qquad (4.1)$$

Z druhého x-ového momentu dostaneme reálnou část druhého kruhového momentu. Pro ostatní momenty platí podobná rovnost, takže kruhové momenty můžeme vyjádřit jako lineární kombinaci plošných momentů a naopak.



Obrázek 4.1: Kružnice přetransformovaná mocninami

Je možné si tedy zvolit kterýkoliv s těchto dvou přístupů a cokoliv lze spočítat z plošných momentů, lze spočítat i z kruhových momentů a naopak. Přesto mají plošné momenty jednu nevýhodu oproti kruhovým momentům. Některé plošné momenty dávají tutéž informaci. Například druhý x-ový moment dává tutéž informaci jako druhý y-ový moment.

$$\int_\Gamma \cos^2\theta \,\mathrm{d}p = \int_\Gamma (1-\sin^2\theta)\mathrm{d}p = 1 - \int_\Gamma \sin^2\theta \,\mathrm{d}p \tag{4.2}$$

Ještě se můžeme zamyslet nad tím, co které momenty vlastně znamenají. Na n-tý moment můžeme pohlížet také tak, že si náhodnou veličinu nejdříve přetransformujeme n-tou mocninou a potom spočítáme střední hodnotu. Pokud umocníme komplexní jednotku na n-tou, vyjde opět komplexní jednotka, ale tentokrát má argument n-krát větší. Jednotková kružnice v komplexní rovině se n-tou mocninou přetransformuje opět na kružnici. Tedy pokud si vezmeme rozdělení na jednotkové kružnici v $\mathbb{C}$ a přetransformujeme ho n-tou mocninou, opět vyjde rozdělení na jednotkové kružnici, ovšem každý bod má n vzorů. Pokud ale v $\mathbb{R}^2$ si vezmeme jednotkovou kružnici a přetransformujeme x-ovou souřadnici n-tou mocninou, vyjde kružnice pouze pro $n = 1$. Pro $n = 2$ vyjde část paraboly a pro $n > 2$ vyjde složitější křivka, dohromady pro sudá n se celá křivka ocitne v kladné části. Pokud si vezmeme rozdělení na jednotkové kružnici v $\mathbb{R}^2$ a souřadnice přetransformujeme mocninami, přičemž aspoň jedna je různá od první mocniny, už nevyjde rozdělení na kružnici.

Druhý kruhový moment může mít uplatnění ve chvíli, kdy zkoumané směry jsou „oboustranné“, například když není rozdíl mezi cestou vedoucí na sever a cestou vedoucí na jih. Pokud zkoumáme, zda-li cesty v nějaké vesnici mají preferovaný směr, brzy narazíme na problém. Pokud ulice vede severojižním směrem, vede na sever, na jih, nebo oběma směry? Pokud se rozhodneme, že směr ulice může být jenom v rozmezí od $-90°$ do $90°$, potom v každé vesnici najdeme preferovaný směr. Tento směr by se navíc změnil, kdybychom

řekli, že směr ulice bude v jiném rozmezí. Pokud budeme používat oba směry najednou, první kruhový moment vyjde vždy nulový. Pokud se budeme rozhodovat náhodně, výsledek závisí na tom, pro kterou ulici byl zvolen který směr. Nic není ale ztraceno, můžeme použít druhý kruhový moment. Druhá mocnina zobrazí opačné směry do jednoho, takže potom nezáleží na tom, jestli se rozhodneme pro jeden směr, nebo směr opačný. Odmocninou z druhého momentu dostaneme preferovaný směr ulic v dané vesnici.

Rozdělení na kružnici se neuplatňuje pouze při zkoumání směrů, ale i ve chvíli, kdy zkoumáme periodické děje a není rozdíl mezi začátkem a koncem periody. Podobnými úvahami můžeme dojít k tomu, že se vyplatí zobrazit si periodický děj na kružnici.

Kapitola 5

Nástin řešení zkoumaných úloh

V této kapitole je popsáno, jak je možné převést zkoumané úlohy na odhad pravděpodobnostního rozdělení. Samotným odhadem pravděpodobnostního rozdělení se zabývají následující kapitoly.

5.1 Úloha s dveřmi

Pro pozitivní a negativní měření uděláme zvlášť model. Odhadneme tedy $p(\vartheta|0)$ a $p(\vartheta|1)$. Podmíněnou pravděpodobnost, že dveře jsou v daný čas otevřené, tedy $P(1|\vartheta)$, vypočítáme pomocí Bayesova vzorce, tedy: $\frac{p(\vartheta|1)P(1)}{p(\vartheta|0)P(0)+p(\vartheta|1)P(1)}$, kde $P(0)$ a $P(1)$ odhadneme pomocí empirického rozdělení. To se zjednoduší na $\frac{p(\vartheta|1)N_1}{p(\vartheta|0)N_0+p(\vartheta|1)N_1}$, kde N_0 je počet měření v negativním modelu a N_1 je počet měření v pozitivním modelu.

5.2 Úloha s vytížeností

Nejdříve si počty hráčů nashlukujeme pomocí EM algoritmu jako směs Gaussových rozdělení. Potom pro každý shluk vytvoříme jeden model. Výsledkem je konvexní kombinace středních hodnot shluků rychlostí, jejíž koeficienty jsou určeny hustotou odhadnutého rozdělení pro jednotlivé shluky. Tento přístup

má tu výhodu, že nemůže dojít k překmitávání, protože výsledek je shora i zdola omezen středními hodnotami jednotlivých shluků.

5.3 Úloha s chodci

Nemá cenu odhadovat, kde se který chodec bude zrovna vyskytovat. Stačí, když budeme pracovat s hustotou pravděpodobnosti, že se nějaký chodec vyskytne na daném místě v daném čase. Tuto hustotu se snažíme odhadnout.

Pravděpodobnost výskytu chodce v daném časoprostorovém okně odhadneme hustotou uprostřed okna vynásobenou velikostí okna. Víme, kolik chodců jsme naměřili, a tedy počet chodců v daném okně se řídí binomickým rozdělením. Vypočítáme střední hodnotu binomického rozdělení, což je pravděpodobnost toho, že se chodec v daném okně vyskytne, vynásobená celkovým počtem chodců.

Kapitola 6

Použitá rozdělení na kružnici

6.1 Von Misesovo rozdělení

Von Misesovo rozdělení je rozdělení na kružnici, dané hustotou pravděpodobnosti [11]:

$$p(\theta|\mu,\kappa) = \frac{\mathrm{e}^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)} \tag{6.1}$$

kde I_0 je nultá hyperbolická Besselova funkce prvního druhu. Von Misesovo rozdělení se podobá normálnímu rozdělení, podstatný rozdíl je v tom, že místo druhé mocniny v exponentu je kosinus. Reálný parametr μ udává směr střední hodnoty a modus, parametr κ, říká, jak moc je rozdělení koncentrované.

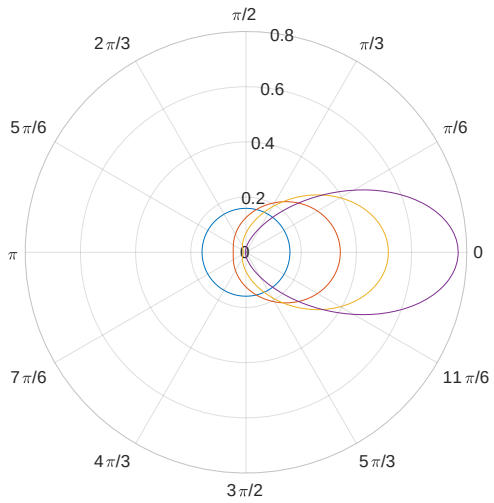
U von Misesova rozdělení známe všechny jeho momenty, jeho n-tý moment je roven [1]:

$$m_n = \mathrm{e}^{n\mu\,\mathrm{i}}\,\frac{I_n(\kappa)}{I_0(\kappa)} \tag{6.2}$$

kde I_n je n-tá hyperbolická Besselova funkce prvního druhu.

Maximálně věrohodný odhad parametru κ je stejný jako odhad pomocí metody momentů, tj. řešení rovnice [14]:

$$|\widehat{m}_1| = \frac{I_1(\kappa)}{I_0(\kappa)} \tag{6.3}$$



Obrázek 6.1: Hustota von Misesova rozdělení pro $\mu = 0$ a $\kappa \in \{0; 1; 2; 4\}$ v polárních souřadnicích. Argumentem je úhel a hustota je znázorněna vzdáleností od počátku.

kde $\widehat{m}_1$ je odhad prvního kruhového momentu, tedy:

$$\widehat{m}_1 = \sum_{n=1}^{N} \vartheta_n \tag{6.4}$$

Funkce $\kappa \mapsto I_1(\kappa)/I_0(\kappa)$ je ryze rostoucí, takže rovnici můžeme řešit metodou bisekce. Newtonovu metodu nepoužíváme, protože tato funkce má malou derivaci pro velké κ. Maximálně věrohodný odhad parametru μ je argument odhadu prvního momentu $\arg(\widehat{m}_1)$.

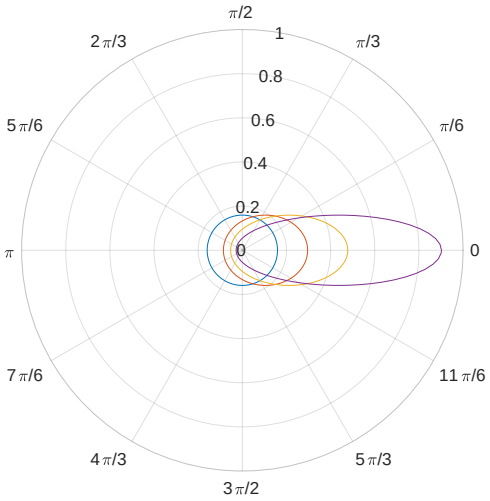
6.2 Kruhové Cauchyovo rozdělení

Kruhové Cauchyovo rozdělení je rozdělení, jehož hustota pravděpodobnosti má tvar [12]:

$$f(\vartheta|\mu) = \frac{1 - |\mu|^2}{2\pi} \cdot \frac{1}{|\vartheta - \mu|^2} \tag{6.5}$$

kde μ je komplexní parametr uvnitř jednotkové kružnice a ϑ je bod na jednotkové kružnici. Pokud chceme, aby rozdělení bylo pro interval $(-\pi; \pi\rangle$, vyjde nám:

$$f(\theta|\mu) = \frac{1 - |\mu|^2}{2\pi} \cdot \frac{1}{|\,\mathrm{e}^{\theta i} - \mu|^2} \tag{6.6}$$



Obrázek 6.2: Hustota kruhového Cauchyova rozdělení pro $\mu \in \{0; 0.3; 0.5; 0.7\}$ v polárních souřadnicích. Argumentem je úhel a hustota je znázorněna vzdáleností od počátku.

kde μ je komplexní parametr a θ je v intervalu $(-\pi; \pi\rangle$. Kruhové momenty tohoto rozdělení jsou:

$$m_n = \mu^n \tag{6.7}$$

Kruhové Cauchyovo rozdělení má tu výhodu, že všechny momenty známe a jdou vyjádřit pomocí primitivních funkcí. U netriviálních rozdělení na kružnici je taková vlastnost spíše výjimečná. Navíc první moment se přímo rovná parametru, takže odhad pomocí metody momentů je zde velmi snadný.

6.3 Odhad parametrů

Parametry kruhového Cauchyova rozdělení můžeme odhadovat pomocí metody momentů nebo pomocí metody maximální věrohodnosti. Odhad metodou momentů je roven:

$$\hat{\mu} = \widehat{m}_1 \tag{6.8}$$

kde $\widehat{m}_1$ je odhad prvního kruhového momentu, tedy:

$$\widehat{m}_1 = \sum_{n=1}^{N} \vartheta_n \tag{6.9}$$

Maximálně věrohodný odhad jde vypočítat pomocí iterativního vzorce [7]:

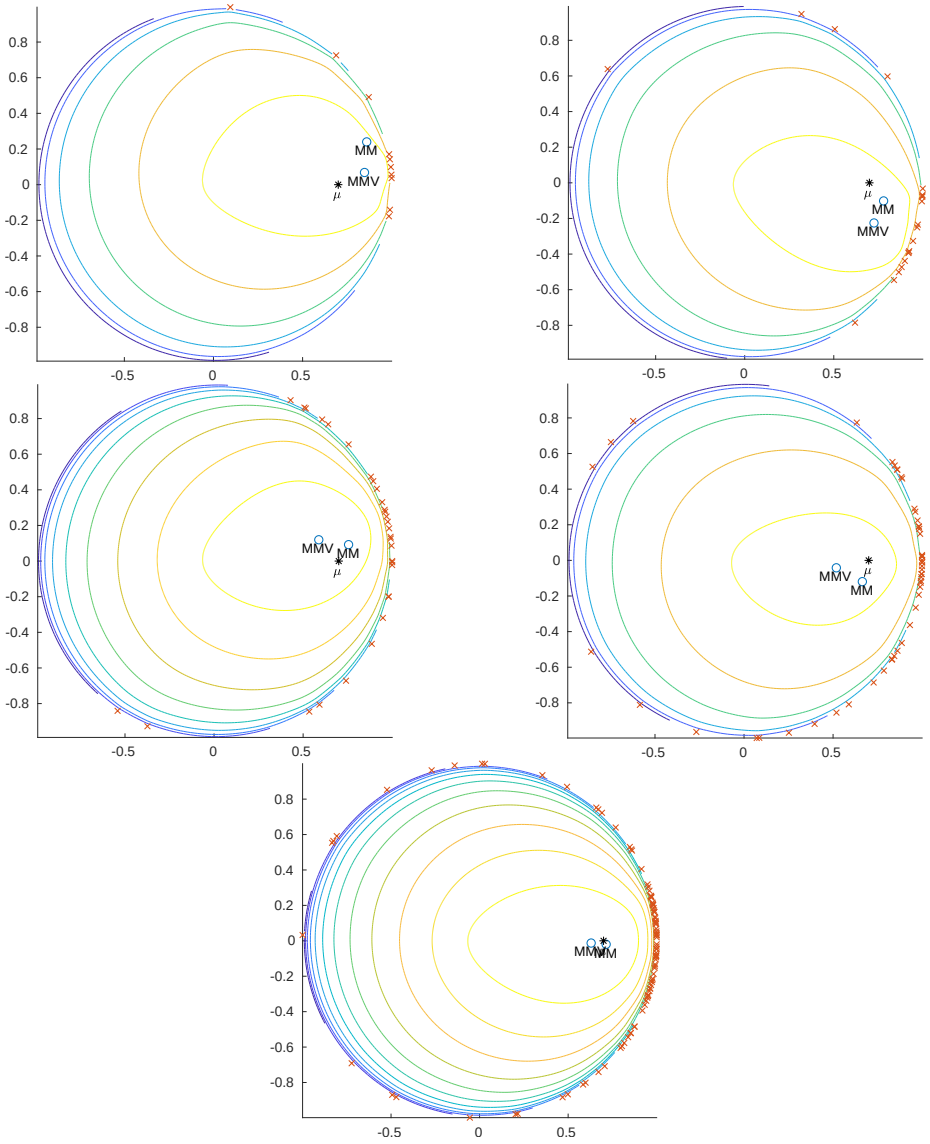
$$\hat{\mu}_0 = 0 \tag{6.10}$$

$$\hat{\mu}_{n+1} = U\left(\frac{1}{N}\sum_{i=1}^{N} U\left(\vartheta_i, \hat{\mu}_n\right), -\hat{\mu}_n\right) \tag{6.11}$$

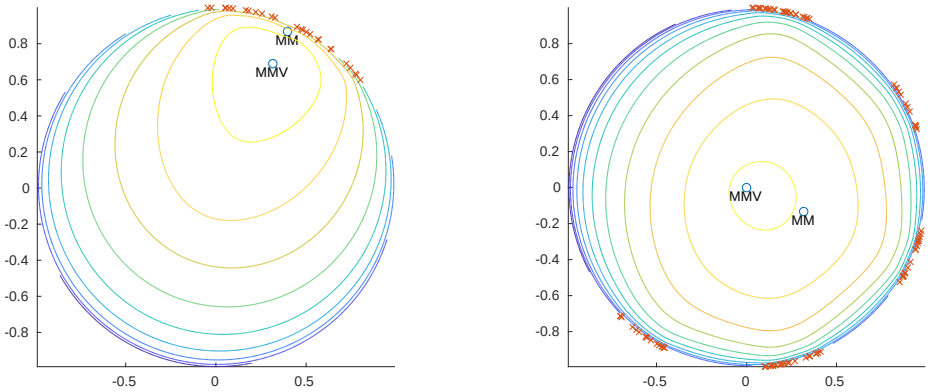
$$U(z, \phi) = \frac{z - \phi}{1 - \bar{\phi} z} \tag{6.12}$$

Zde je několik příkladů, kde byly vygenerovány body podle kruhového Cauchyova a následně byl z těchto bodů odhadnut parametr μ. Barevné křivky ukazují vrstevnice logaritmu věrohodnosti. Hvězdičkou je vyznačena skutečná hodnota μ.

Pokud se snažíme odhadovat data, která pocházejí z jiného rozdělení, odhady se mohou značně lišit. Zvláštní je, že odhad metodou maximální věrohodnosti může vyjít nulový i v případě, že data nejsou symetrická. Data jsou pouze ilustrační, vznikla úpravou dat z [9].



Obrázek 6.3: Odhad parametru μ z 10, 20, 30, 50 a 100 bodů.



Obrázek 6.4: Odhad parametru μ z ilustračních dat

Kapitola 7

Odhad rozdělení na kružnici pomocí metody momentů (MM-vM a MM-kC)

Odhad pomocí metody momentů spočívá v řešení soustavy nelineárních rovnic:

$$m_l = \widehat{m}_l \tag{7.1}$$

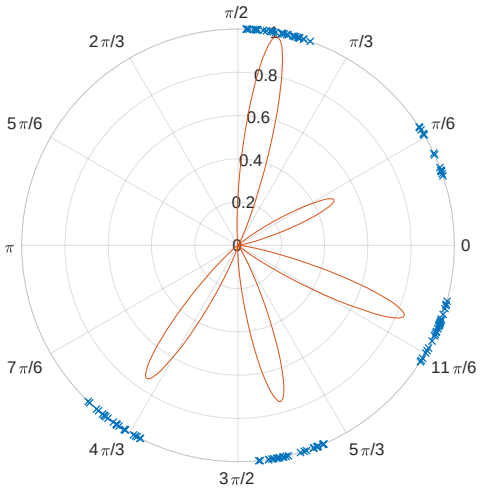
kde m_l je l-tý moment rozdělení a $\widehat{m}_l$ je jeho odhad, což je:

$$\widehat{m}_l = \frac{1}{K} \sum_{k=1}^{K} \vartheta_k^l \tag{7.2}$$

kde ϑ_1 až ϑ_K jsou jednotlivá měření.

Výhoda metody momentů je, že odhad jednotlivých momentů je rychlý a řešení jednotlivých rovnic nezáleží na množství dat. Nevýhodou je, že řešení velkého množství nelineárních rovnic je poměrně obtížný úkol a není proveditelné pro velký počet parametrů. Pro velký počet parametrů je nutné solver restartovat s náhodnými počátečními hodnotami tak dlouho, dokud řešení nenajde, někdy ani to nepomůže.

Pro řešení těchto rovnic byla použita metoda `gsl_multiroot_fsolver_hybrids` z knihovny `gsl` [5].



Obrázek 7.1: Odhad pěti shluků pomocí metody momentů jako směs von Misesových rozdělení. Data jsou pouze ilustrační, byla vytvořena z dat [9].

7.1 Metoda momentů a směs von Misesových rozdělení (MM-vM)

Uvažujme směs N von Misesových rozdělení s parametry μ_1 až μ_N, κ_1 až κ_N a vahami w_1 až w_N. Máme K měření ϑ_1 až ϑ_K. l-tý moment směsi von Misesových rozdělení je roven:

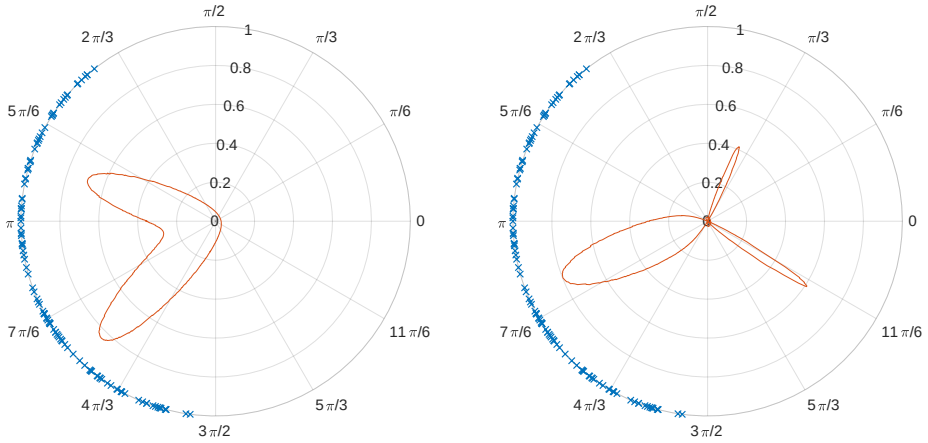
$$m_l = \sum_{n=1}^{N} w_n \, \mathrm{e}^{l\mu_n \, \mathrm{i}} \, \frac{I_l(\kappa_n)}{I_0(\kappa_n)} \tag{7.3}$$

kde I_m je m-tá hyperbolická Besselova funkce prvního druhu. Jelikož parametry von Misesova rozdělení jsou reálná čísla, soustava komplexních rovnic se rozpadne na soustavu reálných rovnic:

$$\Re\widehat{m}_l = \sum_{n=1}^{N} w_n \frac{I_l(\kappa_n)}{I_0(\kappa_n)} \cos(l\mu_n) \tag{7.4}$$

$$\Im\widehat{m}_l = \sum_{n=1}^{N} w_n \frac{I_l(\kappa_n)}{I_0(\kappa_n)} \sin(l\mu_n) \tag{7.5}$$

K odhadu směsi N von Misesových rozdělení je potřeba $\lceil \frac{3}{2}N \rceil$ momentů. Pokud je N liché, tak u odhadu nejvyššího momentu neuvažujeme rovnici odpovídající reálné, nebo imaginární složce, abychom měli stejný počet rovnic a neznámých. Tyto rovnice je možné řešit pouze numericky. Aby numerické řešení lépe konvergovalo, použijeme vhodnou substituci. Potřebujeme, aby solver neuvažoval nepřípustné hodnoty. Pokud máme omezení, že výsledek musí být kladný, tak můžeme například použít druhou mocninu, která způsobí,



Obrázek 7.2: Odhad pěti shluků pomocí metody momentů jako směs dvou (vlevo) a třech (vpravo) kruhových Cauchyových rozdělení. Na obrázku vpravo vidíme, že model sice danou soustavu řeší, ale neodpovídá datům. Data jsou pouze ilustrační, byla vytvořena z dat [9].

že záporná čísla se zobrazí na kladná, a tedy výsledek nemůže vyjít záporný. Toto omezení máme pro κ_n a w_n. Vysoká hodnota κ_n způsobí, že hyperbolická Besselova funkce přeteče. Abychom se tomu pokud možno vyhnuli, použijeme logaritmus k tomu, aby hodnoty κ_n byly navyšovány pomaleji. A protože Besselovy funkce vysokého řádu mají kolem nuly malou derivaci, může solver skončit chybou, když se dostane moc blízko k nule. To vyřešíme přičtením konstanty κ_{min}, kterou musíme předem odhadnout jiným způsobem. Zde ji určujeme metodou pokusu a omylu. Výsledkem těchto úvah je tato substituce:

$$\kappa_n = \ln(\check{\kappa}_n^2 + 1) + \kappa_{min} \tag{7.6}$$

$$w_n = \sqrt{\check{w}_n^2 + 1} - 1 \tag{7.7}$$

Nové proměnné jsou $\check{\kappa}_1$ až $\check{\kappa}_N$, $\check{w}_1$ až $\check{w}_N$. Hodnotu κ_{min} zvolíme jako náš spodní odhad pro parametry κ_1 až κ_n. Pokud aspoň tušíme, v jakém intervalu by se mohlo κ_{min} nacházet, tak v každém pokusu o vyřešení soustavy dosadíme náhodnou hodnotu v tomto intervalu. Pokud vůbec žádný spodní odhad nemáme, dosadíme nulu. Touto substitucí se zbavíme všech omezujících podmínek pro neznámé, kromě té, že součet vah musí být roven jedné. V praxi se ukazuje, že pokud metoda konverguje, dává smysluplný odhad.

7.2 Metoda momentů a směs kruhových Cauchyových rozdělení (MM-kC)

Uvažujme směs N kruhových Cauchyových rozdělení s parametry μ_1 až μ_N a vahami w_1 až w_N. Máme K měření ϑ_1 až ϑ_K. l-tý moment směsi kruhových

Cauchyových rozdělení je roven:

$$m_l = \sum_{n=1}^{N} w_n \mu_n^l \tag{7.8}$$

Sice jsou všechny rovnice polynomiální, ale řešení je mnoho a je těžké najít to správné. Navíc metoda momentů nám neposkytuje nástroj, pomocí kterého bychom si mohli z množiny kořenů vybrat ten správný. Můžeme pouze zkontrolovat, jestli se skutečně jedná o pravděpodobnostní rozdělení, a pokud ne, tak se pokusíme najít jiné řešení. Ani tak ale není zaručeno, že nalezený model odpovídá datům. V praxi se ukazuje, že často najdeme nějaké špatné řešení.

Kapitola 8

Odhad rozdělení na kružnici pomocí EM algoritmu (EM-vM a EM-kC)

Metoda zde označená jako EM-vM modeluje pravděpodobnostní rozdělení jako směs von Misesových rozdělení a metoda zde označená jako EM-kC modeluje pravděpodobnostní rozdělení jako směs kruhových Cauchyových rozdělení.

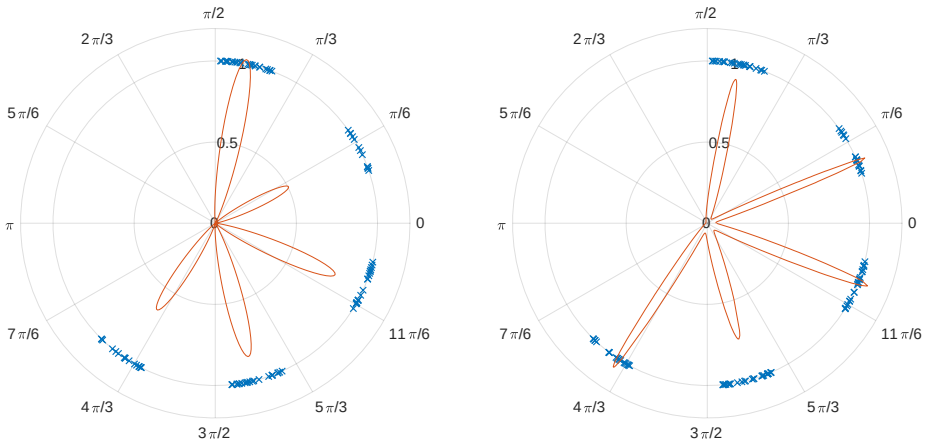
8.1 EM algoritmus pro směs von Misesových rozdělení (EM-vM)

Uvažujme směs N von Misesových rozdělení s reálnými parametry μ_1 až μ_N, κ_1 až κ_N a vahami w_1 až w_N. Máme K měření ϑ_1 až ϑ_K na jednotkové kružnici. Parametry této směsi můžeme odhadnout pomocí EM algoritmu, který probíhá následujícím způsobem.

1. Inicializujeme parametry μ_1 až μ_N, κ_1 až κ_N a váhy w_1 až w_N náhodně tak, aby $\mu_n \in (-\pi; \pi\rangle$, $\kappa_n > 0$ a $\sum_{n=1}^{N} w_n = 1$. Potom opakujeme následující dva kroky:

E: Stanovíme stupně příslušnosti:

$$\alpha_{k,n} = \frac{w_n f(\vartheta_k|\mu_n, \kappa_n)}{\sum_{i=1}^{N} w_i f(\vartheta_k|\mu_i, \kappa_i)} \tag{8.1}$$



Obrázek 8.1: Odhad pěti shluků pomocí EM algoritmu jako směs von Misesových rozdělení (vlevo) a jako směs kruhových Cauchyových rozdělení (vpravo). Data jsou pouze ilustrační, byla vytvořena z dat [9].

M: Aktualizujeme váhy:

$$w_n = \frac{1}{K} \sum_{k=1}^{K} \alpha_{k,n} \tag{8.2}$$

a odhadneme parametry jako maximální hodnotu logaritmu věrohodnostní funkce. Pro von Misesovo rozdělení je to řešení rovnice [14]:

$$\left| \frac{\sum_{k=1}^{K} \alpha_{k,n} \vartheta_k}{K w_n} \right| = \frac{I_1(\kappa_n)}{I_0(\kappa_n)} \tag{8.3}$$

$$\arg\left(\sum_{k=1}^{K} \alpha_{k,n} \vartheta_k \right) = \mu_n \tag{8.4}$$

Pro zlepšení numerické stability je dobré následně zkontrolovat, jestli hodnota κ_n není moc velká, a pokud přeroste určitou mez, tak ji ořízneme. V této práci to řešíme tak, že rovnici řešíme bisekcí a jako počáteční interval nastavíme $\langle 0; 300 \rangle$ a pokud se ukáže, že řešení není v tomto intervalu, prohlásíme horní mez za řešení.

Kroky E a M opakujeme tak dlouho, dokud věrohodnost roste dostatečně rychle. Můžeme zvolit i jiné ukončovací kritérium.

8.2 EM algoritmus pro směs kruhových Cauchyových rozdělení (EM-kC)

Uvažujme směs N kruhových Cauchyových rozdělení s komplexními parametry μ_1 až μ_N a vahami w_1 až w_N. Máme K měření ϑ_1 až ϑ_K na jednotkové kružnici. Parametry této směsi můžeme odhadnout pomocí EM algoritmu, který probíhá následujícím způsobem.

1. Inicializujeme parametry μ_1 až μ_N a váhy w_1 až w_N náhodně tak, aby $|\mu_n| < R_{max} < 1$ a $\sum_{n=1}^{N} w_n = 1$. Potom opakujeme následující dva kroky:

E: Stanovíme stupně příslušnosti:

$$\alpha_{k,n} = \frac{w_n f(\vartheta_k|\mu_n)}{\sum_{i=1}^{N} w_i f(\vartheta_k|\mu_i)} \tag{8.5}$$

M: Aktualizujeme váhy:

$$w_n = \frac{1}{K}\sum_{k=1}^{K} \alpha_{k,n} \tag{8.6}$$

a odhadneme parametry pomocí maxima logaritmu věrohodnostní funkce. Pro kruhové Cauchyovo rozdělení je to limita posloupnosti $\left(\mu_n^{(l)}\right)_{l=0}^{\infty}$ dané rekurentně [7]:

$$\mu_n^{(0)} = 0 \tag{8.7}$$

$$\mu_n^{(l+1)} = U\left(\frac{\sum_{k=1}^{K} \alpha_{k,n} U\left(\vartheta_k, \mu_n^{(l)}\right)}{K w_n}, -\mu_n^{(l)}\right) \tag{8.8}$$

$$U(z,\phi) = \frac{z-\phi}{1-\bar{\phi}z} \tag{8.9}$$

Pro zlepšení numerické stability je dobré následně zkontrolovat absolutní hodnotu μ_n, a pokud je moc velká (tj. μ_n je moc blízko jednotkové kružnici), tak μ_n přeškálujeme. V této práci sledujeme, jestli $|\mu_n|$ není větší než $R_{max} = 0.999$. Pokud je větší, tak μ_n přeškálujeme tak, aby $|\mu_n| = R_{max}$.

K těmto vzorcům dojdeme snadno logickou úvahou, když vyjdeme z maximálně věrohodného odhadu pro daná rozdělení. Kdyby některé ϑ_i bylo naměřeno dvakrát, můžeme si výpočet zkrátit tím, že odpovídající sčítanec vynásobíme dvěma. Kdyby bylo naměřeno půlkrát, odpovídající sčítanec vynásobíme jednou polovinou. A kdyby bylo naměřeno $\alpha_{k,n}$-krát, tak odpovídající

sčítanec vynásobíme koeficientem $\alpha_{k,n}$. A celou sumu musíme vydělit počtem měření, což je v tomto případě $\sum_{k=1}^{K} \alpha_{k,n} = K w_n$.

Kroky E a M opakujeme tak dlouho, dokud věrohodnost roste dostatečně rychle. Můžeme zvolit i jiné ukončovací kritérium.

Kapitola 9

Logistická regrese pro von Misesovo rozdělení

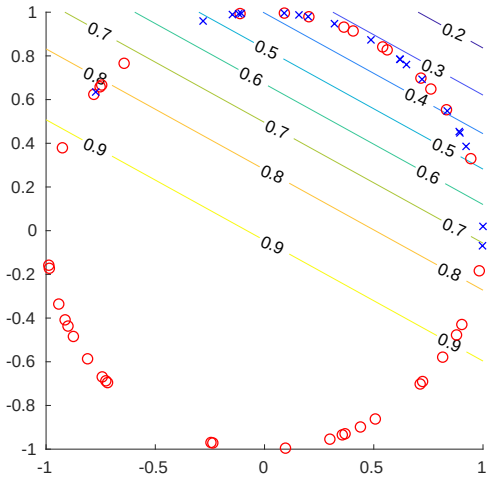
Uvažujme dvě třídy, přičemž každá má von Misesovo rozdělení, ale s jinými parametry. První třída má parametry μ_1 a κ_1, druhá třída má parametry μ_2 a κ_2. Aposteriorní pravděpodobnost první třídy je potom:

$$
\begin{aligned}
P(1|\theta) &= \frac{P(1)\,p(\theta|1)}{P(1)\,p(\theta|1) + P(2)\,p(\theta|2)} = \\
&= \frac{P(1)\frac{\exp(\kappa_1\cos(\mu_1-\theta))}{2\pi I_0(\kappa_1)}}{P(1)\frac{\exp(\kappa_1\cos(\mu_1-\theta))}{2\pi I_0(\kappa_1)} + P(2)\frac{\exp(\kappa_2\cos(\mu_2-\theta))}{2\pi I_0(\kappa_2)}} = \\
&= \frac{1}{1 + \frac{P(2)I_0(\kappa_2)}{P(1)I_0(\kappa_1)}\exp\left(\kappa_2\cos(\mu_2-\theta) - \kappa_1\cos(\mu_1-\theta)\right)} = \\
&= \frac{1}{1 + \frac{P(1)I_0(\kappa_1)}{P(2)I_0(\kappa_2)}\mathrm{e}^{\kappa_2\cos\mu_2\cos\theta+\kappa_2\sin\mu_2\sin\theta-\kappa_1\cos\mu_1\cos\theta-\kappa_1\sin\mu_1\sin\theta}} = \\
&= \frac{1}{1 + \frac{P(1)I_0(\kappa_1)}{P(2)I_0(\kappa_2)}\mathrm{e}^{\cos\theta(\kappa_2\cos\mu_2-\kappa_1\cos\mu_1)+\sin\theta(\kappa_2\sin\mu_2-\kappa_1\sin\mu_1)}} = \\
&= \frac{1}{1 + \mathrm{e}^{\ln(P(1)I_0(\kappa_1))-\ln(P(2)I_0(\kappa_2))+\cos\theta(\kappa_2\cos\mu_2-\kappa_1\cos\mu_1)+\sin\theta(\kappa_2\sin\mu_2-\kappa_1\sin\mu_1)}}
\end{aligned}
$$

Nyní definujeme:

$$
\breve{\theta} = \begin{bmatrix} 1 \\ \cos\theta \\ \sin\theta \end{bmatrix} \tag{9.1}
$$

$$
w = \begin{bmatrix} \ln(I_0(\kappa_1)) + \ln P(1) - \ln(I_0(\kappa_2)) - \ln P(2) \\ \kappa_2\cos\mu_2 - \kappa_1\cos\mu_1 \\ \kappa_2\sin\mu_2 - \kappa_1\sin\mu_1 \end{bmatrix} \tag{9.2}
$$



Obrázek 9.1: Příklad logistické regrese. Červené kroužky značí měření první třídy, modré křížky značí měření druhé třídy. Vrstevnice vyjadřují odhad aposteriorní pravděpodobnosti první třídy. Data jsou pouze ilustrační.

Aposteriorní pravděpodobnost první třídy tedy můžeme napsat jako:

$$p(1|\theta) = \frac{1}{1 + \mathrm{e}^{\langle w, \check{\theta} \rangle}} \tag{9.3}$$

To odpovídá logistické regresi.

Když máme dva shluky s vícerozměrným Gaussovým rozdělením s podobnými kovariančními maticemi, můžeme aposteriorní pravděpodobnost počítat pomocí logistické regrese. Ale v případě, kdy jsou kovarianční matice výrazně odlišné, musíme před využitím samotné logistické regrese provést tzv. dimensionality lifting. U von Misesova rozdělení je situace jiná, právě jsme ukázali, že logistickou regresi můžeme použít pro libovolné parametry jednotlivých shluků, aniž bychom potřebovali dimensionality lifting.

Pokud budeme tvrdit, že von Misesovo rozdělení je rozdělení na intervalu $(-\pi; \pi\rangle$, tak to pro nás znamená, že zobrazení na jednotkovou kružnici je vhodným dimensionality liftingem pro von Misesovo rozdělení.

9.1 Logistická regrese pro von Misesovo-Fisherovo rozdělení

Podobná věta platí i pro obecnější, von Misesovo-Fisherovo rozdělení, což je zobecnění von Misesova rozdělení pro vícerozměrnou sféru. Jeho hustota pro

p rozměrů je definovaná jako [15] :

$$f(x|\mu,\kappa) = c_p(\kappa)\,\mathrm{e}^{\kappa\langle\mu,x\rangle} \tag{9.4}$$

$$c_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)} \tag{9.5}$$

kde I_s značí hyperbolickou Besselovu funkci prvního druhu. Parametry musí splňovat: $\mu \in \mathbb{R}^p$, $||\mu|| = 1$, $\kappa \in \mathbb{R}$, $\kappa \geq 0$. A protože se jedná o rozdělení na sféře, tak: $x \in \mathbb{R}^p$, $||x|| = 1$. Podobně jako v předchozím případě můžeme vypočítat aposteriorní pravděpodobnost první třídy:

$$\begin{aligned}
P(1|x) &= \frac{P(1)\,p(\theta|1)}{P(1)\,p(\theta|1) + P(2)\,p(\theta|2)} = \\
&= \frac{P(1)c_p(\kappa_1)\,\mathrm{e}^{\kappa_1\langle\mu_1|x\rangle}}{P(1)c_p(\kappa_1)\,\mathrm{e}^{\kappa_1\langle\mu_1|x\rangle} + P(2)c_p(\kappa_2)\,\mathrm{e}^{\kappa_2\langle\mu_2|x\rangle}} = \\
&= \frac{\mathrm{e}^{\ln(P(1)c_p(\kappa_1))+\kappa_1\langle\mu_1|x\rangle}}{\mathrm{e}^{\ln(P(1)c_p(\kappa_1))+\kappa_1\langle\mu_1|x\rangle} + \mathrm{e}^{\ln(P(2)c_p(\kappa_2))+\kappa_2\langle\mu_2|x\rangle}} = \\
&= \frac{1}{1 + \mathrm{e}^{\ln(P(2)c_p(\kappa_2))-\ln(P(1)c_p(\kappa_1))+\kappa_2\langle\mu_2|x\rangle-\kappa_1\langle\mu_1|x\rangle}} = \\
&= \frac{1}{1 + \mathrm{e}^{\ln(P(2)c_p(\kappa_2))-\ln(P(1)c_p(\kappa_1))+\langle\kappa_2\mu_2-\kappa_1\mu_1|x\rangle}}
\end{aligned} \tag{9.6}$$

Nyní definujeme

$$\check{x} = \begin{bmatrix} 1 \\ x \end{bmatrix} \tag{9.7}$$

$$w = \begin{bmatrix} \ln(P(2)) + \ln(c_p(\kappa_2)) - \ln(P(1)) - \ln(c_p(\kappa_1)) \\ \kappa_2\mu_2 - \kappa_1\mu_1 \end{bmatrix} \tag{9.8}$$

Aposteriorní pravděpodobnost první třídy tedy můžeme napsat jako:

$$p(1|\theta) = \frac{1}{1 + \mathrm{e}^{\langle w,\check{x}\rangle}} \tag{9.9}$$

To odpovídá logistické regresi.

Podobně jako v předchozím případě, logistickou regresi můžeme použít pro libovolné dva shluky s von Misesovým-Fisherovým rozdělením. Nevýhodou von Misesova-Fisherova rozdělení, alespoň podle výše uvedené definice, je to, že von Misesovo-Fisherovo rozdělení je osově souměrné podle osy μ, takže nemůže modelovat data, která nejsou osově souměrná. Pokud bychom přidali další parametry tak, aby rozdělení mohlo být v některé ose protáhlejší, situace by se zkomplikovala. Už by nešlo tvrdit, že logistickou regresi lze použít pro shluky s libovolnými parametry.

Kapitola 10

Adaboost$_{\mathbb{R}}$ [13]

Logistická regrese nabízí pouze jednoduchou klasifikaci, nestačí však na řešení zkoumaných úloh. Nabízí se metoda Adaboost$_{\mathbb{R}}$ [13], což je zobecnění metody Adaboost pro spojité klasifikátory. Máme trénovací sadu o N prvcích, máme měření $y_1, \ldots, y_N \in \{-1; 1\}$ a jejich časy $x_1, \ldots, x_N \in \mathbb{R}$. Na testovací sadu aplikujeme metodu FreMEn [8], abychom zjistili periodicity. Pomocí metody FreMEn [8] si zjistíme M nejvýraznějších periodicit v trénovací množině, označíme je T_1 až T_M. Nastavíme počáteční váhy rovnoměrně $w_{1,1}, \ldots, w_{1,N} = \frac{1}{N}$. Pro $t \in 1, \ldots, I$, kde I je předem zvolený počet iterací, provedeme:

Pro každou periodu T_1 až T_M zobrazíme body x_1 až x_N následovně:

$$D_j(x_i) = \begin{bmatrix} \cos\left(2\pi x_i / T_j\right) \\ \sin\left(2\pi x_i / T_j\right) \\ 1 \end{bmatrix} \tag{10.1}$$

$$\tag{10.2}$$

pro $i = 1, \ldots, M$ a ještě:

$$D_{M+1}(x_i) = \begin{bmatrix} \frac{t - \frac{1}{2}(x_{max} + x_{min})}{\frac{1}{2}(x_{max} - x_{min})} \\ \left(\frac{t - \frac{1}{2}(x_{max} + x_{min})}{\frac{1}{2}(x_{max} - x_{min})}\right)^2 \\ 1 \end{bmatrix} \tag{10.3}$$

$$x_{min} = \min_i x_i \tag{10.4}$$

$$x_{max} = \max_i x_i \tag{10.5}$$

Pro každé zobrazení D_1 až D_{M+1} natrénujeme logistickou regresi. V případě zobrazení D_1 až D_M jde o logistickou regresi pro von Misesovo rozdělení, v případě zobrazení D_{M+1} jde o logistickou regresi pro normální rozdělení. Postup je však stejný. Logistickou regresi spočítáme jako minimum funkce:

$$l_{t,j}(u_{t,j}) = \frac{\sum_{i=1}^{N} w_{t,i} \ln\left(1 + \mathrm{e}^{-y_i \langle u_{t,j}, D_j(x_i)\rangle}\right)}{\sum_{i=1}^{N} w_{t,i}} \tag{10.6}$$

To spočítáme pomocí gradientního sestupu. Zvolíme si počáteční odhad $u_{t,j,0}$ a délku kroku k_0. Stanovíme ϵ podle toho, jak přesného výsledku chceme dosáhnout, například 10^{-3}. Spočítáme gradient $\nabla l_{t,j}(u_{t,j,0})$ přes vektor u. V každém kroku spočítáme: Pokud $l_{t,j}(u_{t,j,s} - k_i \nabla l_{t,j}(u_{t,j,0})) < l_{t,j}(u_{t,j,s})$, tak:

$$u_{t,j,s+1} := u_{t,j,s} - k_i \nabla(l_{t,j}(u_{t,j,0})) \tag{10.7}$$

$$k_{s+1} := 2k_s \tag{10.8}$$

jinak:

$$u_{t,j,s+1} := u_{t,j,s} \tag{10.9}$$

$$k_{s+1} := \frac{1}{2} k_s \tag{10.10}$$

Pokud $||u_{t,j,i+s} - u_{t,j,s}|| < \epsilon$, skončíme a vrátíme $u_{t,j}$. Pokud dojde k přetečení datového typu u libovolné proměnné, vrátíme poslední hodnotu $u_{t,j}$, ve které ještě k přetečení nedošlo. Výsledná klasifikace pro zobrazení D_j je rovna:

$$h_{t,j}(x) = 2\frac{1}{1 + \mathrm{e}^{-\langle u_{t,j}, D_j(x)\rangle}} - 1 \tag{10.11}$$

Z těchto klasifikátorů vybereme ten, pro který $l_{t,j}(u_{t,j})$ je nejmenší. Označíme ho $h_t(x)$.

Spočítáme:

$$\mu_t = \sum_{i=1}^{N} w_{t,i} y_i h_t(x_i) \tag{10.12}$$

$$\alpha_t = \frac{1}{2} \ln \frac{1 + \mu_t}{1 - \mu_t} \tag{10.13}$$

$$w_{t+1,i} = w_{t,i} \frac{1 - \mu_t y_i h_t(x_i)}{1 - \mu_t^2} \tag{10.14}$$

$$\tag{10.15}$$

Výsledný klasifikátor je potom:

$$H(x) = \frac{\sum_{i=1}^{I} \alpha_i h_i(x)}{\sum_{i=1}^{I} \alpha_i} \tag{10.16}$$

Kapitola 11

Neuronové sítě (NS)

Neuronové sítě jsou proslulé tím, že fungují jako černá krabička, která záhadně řeší nejrůznější úlohy. Tady tomu tak nebude, ukážeme, že naše neuronová síť je založená na směrové statistice. I když se bude navenek jevit jako neuronová síť jako každá jiná, uvnitř se budou řešit skutečné statistické úlohy a bude dokonce možné výsledky interpretovat pomocí fuzzy logiky. Adaboost umí pouze sečíst jednotlivé slabé klasifikátory s různými vahami. Naše neuronová síť s nimi umí provádět i jiné fuzzy logické operace. Než začneme stavět neuronovou síť a vysvětlovat její vlastnosti, definujeme jednotlivé její vrstvy.

Vrstva dimensionality-liftingu má jako vstup jedno reálné číslo t a jako výstup má N (případně $N-1$) trojprvkových vektorů $y_1, \ldots, y_N$, přičemž prvními dvěma prvky jsou zobrazení na kružnici nebo parabolu a třetím prvkem je jednička. Tato vrstva zobrazuje čas na kružnice podle různých period $T_1, \ldots, T_{N-1}$, které předem stanovíme pomocí Fourierovy transformace, tedy:

$$y_i = \begin{bmatrix} \cos\left(2\pi t/T_i\right) \\ \sin\left(2\pi t/T_i\right) \\ 1 \end{bmatrix} \tag{11.1}$$

pro $i = 1, \ldots, N-1$. Také zobrazí čas na parabolu, přičemž parametry tohoto zobrazení jsou předem stanoveny tak, aby celá trénovací množina se těsně zobrazila do mocniny intervalu $\langle -1; 1 \rangle$, tedy:

$$y_N = \begin{bmatrix} \frac{t-\frac{1}{2}(t_{max}+t_{min})}{\frac{1}{2}(t_{max}-t_{min})} \\ \left(\frac{t-\frac{1}{2}(t_{max}+t_{min})}{\frac{1}{2}(t_{max}-t_{min})}\right)^2 \\ 1 \end{bmatrix} \tag{11.2}$$

kde t_{max} je největší čas v trénovací sadě a t_{min} je nejmenší čas v trénovací sadě. V této vrstvě během učení neměníme žádné parametry. Pokud máme málo dat, zobrazení na parabolu vynecháváme, aby nedošlo k přefitování.

Logistická vrstva se musí nacházet bezprostředně za vrstvou dimensionality-liftingu a na každý trojprvkový vektor x_i z jejího výstupu aplikuje několik logistických funkcí, tedy:

$$y_{i,j} = \frac{1}{1 + \mathrm{e}^{\langle x_i, w_{i,j} \rangle}} \tag{11.3}$$

kde $w_{i,j}$ jsou parametry vrstvy, které měníme během učení. V této vrstvě nemícháme jednotlivé vektory x_i mezi sebou. Výstupem je dlouhý vektor poskládaný z $y_{i,j}$. V této vrstvě se provádí logistická regrese, která plyne z předchozích kapitol. Je sice pravda, že tentýž nápad bychom mohli dostat i bez těchto poznatků, ale potom bychom vůbec nevěděli, co se tu děje.

Lineární vrstva bere jako vstup vektor x libovolné délky a vrací vektor y libovolné délky. Tyto délky se mohou lišit. Výstup je definován jako:

$$y = W \begin{bmatrix} x \\ 1 \end{bmatrix} \tag{11.4}$$

kde W je matice reálných čísel s takovými rozměry, aby vztah dával smysl. Matici W měníme během učení.

Fuzzy logická vrstva je vrstva, která provádí operace minimum a maximum pro každou dvojici prvků vstupního vektoru x. Konkrétně:

$$y_{i,j} = \begin{cases} \max(x_i, x_j) & \text{pro } i < j \\ x_i & \text{pro } i = j \\ \min(x_i, x_j) & \text{pro } i > j \end{cases} \tag{11.5}$$

Operace minimum a maximum slučujeme do jedné vrstvy kvůli snazší implementaci. Kdybychom prováděli jenom jednu z operací, tak by skoro polovina výstupů byla přebytečná. To bychom mohli vyřešit tím, že bychom tu druhou polovinu nepočítali vůbec, ale zase bychom museli implementovat trojúhelníkové pole. Když tyto operace budeme provádět v jedné vrstvě, tak si tím nejen ušetříme čas při implementaci, ale také umožníme neuronové síti, aby si vybrala tu vhodnou operaci. Výstupem je dlouhý vektor, který se skládá z prvků $y_{i,j}$.

Neuronovou síť postavíme tak, že na začátek umístíme vrstvu dimensionality-liftingu a logistickou vrstvu. Potom střídáme lineární vrstvy a fuzzy logické vrstvy. Na konci by měla být lineární vrstva. Výstupem může být skalár nebo vektor podle toho, jakou úlohu řešíme.

11.1 Směrová statistika a fuzzy logika v neuronové síti

Nejdříve si povšimněme, že v naší neuronové síti se provádí logistická regrese, a sice v prvních dvou vrstvách. V rámci těchto dvou vrstev nedokážeme poznat, jaké dvě třídy jsou zde klasifikovány, ale zjevně tu jsou nějaké dvě třídy, které odhadujeme von Misesovým (případně normálním) rozdělením. Také můžeme říct, že se snažíme detekovat nějakou fuzzy množinu, která je daná příslušností do jedné z těchto dvou tříd. Po natrénování sítě si můžeme data z logistické vrstvy přečíst a podívat se, jaké fuzzy množiny mají na rozhodování vliv.

Výsledky dalších vrstev můžeme interpretovat tak, že každý výstup je konvexní kombinací dvou hodnot, přičemž koeficient této kombinace je nějaká fuzzy logická formule, která využívá operace konjunkce, disjunkce, negace a vážený průměr. Vážený průměr není běžnou fuzzy logickou operací, můžeme si ji představit tak, že v přirozeném jazyce bychom řekli výraz pro konjunkci, ale zároveň bychom tím mysleli, že nesplnění jedné podformule moc nevadí. Například ve větě: „Město A je velké právě tehdy, když město A má hodně obyvatel a město A má velkou rozlohu," máme na mysli, že když nějaké město má velkou rozlohu, ale nemá hodně obyvatel, je jenom trochu velké.

Nyní definujeme fuzzy logické operace, které budeme používat. V následujících vzorcích platí, že $\alpha_1, \ldots, \alpha_N$ jsou libovolné fuzzy logické formule a $u(\alpha)$ je ohodnocení formule α. Negaci definujeme jako unární operaci, pro kterou platí:

$$u(\neg\alpha) = 1 - u(\alpha) \tag{11.6}$$

Konjunkci a disjunkci definujeme jako binární operace, pro které platí:

$$u(\alpha_1 \vee \alpha_2) = \max(u(\alpha_1), u(\alpha_2)) \tag{11.7}$$

$$u(\alpha_1 \wedge \alpha_2) = \min(u(\alpha_1), u(\alpha_2)) \tag{11.8}$$

Vážený průměr definujeme jako N-ární logickou operaci operaci ($N \geq 1$ libovolné), pro kterou platí:

$$u\left(\mathrm{Vp}_{w_1,\ldots,w_N}(\alpha_1, \ldots, \alpha_N)\right) = \frac{\sum_{i=1}^{N} w_i u(\alpha_i)}{\sum_{i=1}^{N} w_i} \tag{11.9}$$

kde $w_1, \ldots, w_N \geq 0$ a $\sum_{i=1}^{N} w_i > 0$. Logické konstanty $\mathbf{0}$ a $\mathbf{1}$ definujeme takto:

$$u(\mathbf{0}) = 0 \tag{11.10}$$

$$u(\mathbf{1}) = 1 \tag{11.11}$$

$$\tag{11.12}$$

Fuzzy logickou kvaziformuli definujeme jako vzorec ve tvaru:

$$xu(\alpha) + y \tag{11.13}$$

kde α je fuzzy logická formule, $x \geq 0$, $y \in \mathbb{R}$. Formule α využívá pouze výše vyjmenované operace, logické konstanty **0** a **1** a fuzzy logické proměnné. Operace v lineárních a fuzzy logických vrstvách můžeme popsat pomocí operací: sčítání, násobení nezáporným číslem, změna znaménka, přičtení konstanty, minimum, maximum. Ukážeme si, že pokud na dvou fuzzy logických kvaziformulích provedeme kteroukoliv z těchto operací, výsledek je opět fuzzy logická kvaziformule. Nechť α, β jsou fuzzy logické formule, $x_1, x_2, w \geq 0$, $y_1, y_2 \in \mathbb{R}$

$$\begin{aligned}(x_1u(\alpha) + y_1) + (x_2u(\beta) + y_2) &= (x_1 + x_2)\frac{x_1u(\alpha) + x_2u(\beta)}{x_1 + x_2} + (y_1 + y_2) = \\ &= (x_1 + x_2)u\left(\mathrm{Vp}_{x_1,x_2}(\alpha, \beta)\right) + (y_1 + y_2)\end{aligned} \tag{11.14}$$

$$w(x_1u(\alpha) + y_1) = wx_1u(\alpha) + wy_1 \tag{11.15}$$

$$\begin{aligned}-(x_1u(\alpha) + y_1) &= x_1(1 - u(\alpha)) + (-x_1 - y_1) = \\ &= x_1u(\neg\alpha) + (-x_1 - y_1)\end{aligned} \tag{11.16}$$

$$(x_1u(\alpha) + y_1) + y_2 = x_1u(\alpha) + (y_1 + y_2) \tag{11.17}$$

Bez újmy na obecnosti předpokládejme, že $y_1 \geq y_2$. Potom:

$$\max(x_1u(\alpha) + y_1; x_2u(\beta) + y_2) = \max(x_1u(\alpha) + y_1 - y_2; x_2u(\beta)) + y_2 \tag{11.18}$$

$$\min(x_1u(\alpha) + y_1; x_2u(\beta) + y_2) = \min(x_1u(\alpha) + y_1 - y_2; x_2u(\beta)) + y_2 \tag{11.19}$$

Pokud $x_1 + y_1 - y_2 \geq x_2$:

$$\begin{aligned}&\max(x_1u(\alpha) + y_1 - y_2; x_2u(\beta)) + y_2 = \\ &= (x_1 + y_1 - y_2)\max\left(\frac{x_1u(\alpha) + y_1 - y_2}{x_1 + y_1 - y_2}; \frac{x_2u(\beta)}{x_1 + y_1 - y_2}\right) + y_2\end{aligned} \tag{11.20}$$

$$\begin{aligned}&\min(x_1u(\alpha) + y_1 - y_2; x_2u(\beta)) + y_2 = \\ &= (x_1 + y_1 - y_2)\min\left(\frac{x_1u(\alpha) + y_1 - y_2}{x_1 + y_1 - y_2}; \frac{x_2u(\beta)}{x_1 + y_1 - y_2}\right) + y_2\end{aligned} \tag{11.21}$$

Výrazy uvnitř maxim a minim jsou sice v rozmezí $\langle 0; 1\rangle$, ale abychom je mohli převést na ohodnocení logických formulí, musíme zavést logické konstanty.

Potom už snadno tyto výrazy převedeme na ohodnocení váženého průměru.

$$
\begin{aligned}
&(x_1+y_1-y_2)\max\left(\frac{x_1u(\alpha)+y_1-y_2}{x_1+y_1-y_2};\frac{x_2u(\beta)}{x_1+y_1-y_2}\right)+y_2=\\
&=y_2+(x_1+y_1-y_2)\cdot\\
&\cdot\max\left(\frac{x_1u(\alpha)+(y_1-y_2)u(\mathbf{1})}{x_1+y_1-y_2};\frac{x_2u(\beta)+(x_1+y_1-y_2-x_2)u(\mathbf{0})}{x_1+y_1-y_2}\right)=\\
&=y_2+(x_1+y_1-y_2)\cdot\\
&\cdot\max\left(\frac{x_1u(\alpha)+(y_1-y_2)u(\mathbf{1})}{x_1+(y_1-y_2)};\frac{x_2u(\beta)+(x_1+y_1-y_2-x_2)u(\mathbf{0})}{x_2+(x_1+y_1-y_2-x_2)}\right)=\\
&=(x_1+y_1-y_2)u\left(\mathrm{Vp}_{x_1,y_1-y_2}(\alpha,\mathbf{1})\vee\mathrm{Vp}_{x_2,x_1+y_1-y_2-x_2}(\beta,\mathbf{0})\right)+y_2
\end{aligned}
\tag{11.22}
$$

$$
\begin{aligned}
&(x_1+y_1-y_2)\min\left(\frac{x_1u(\alpha)+y_1-y_2}{x_1+y_1-y_2};\frac{x_2u(\beta)}{x_1+y_1-y_2}\right)+y_2=\\
&=y_2+(x_1+y_1-y_2)\cdot\\
&\cdot\min\left(\frac{x_1u(\alpha)+(y_1-y_2)u(\mathbf{1})}{x_1+y_1-y_2};\frac{x_2u(\beta)+(x_1+y_1-y_2-x_2)u(\mathbf{0})}{x_1+y_1-y_2}\right)=\\
&=y_2+(x_1+y_1-y_2)\cdot\\
&\cdot\min\left(\frac{x_1u(\alpha)+(y_1-y_2)u(\mathbf{1})}{x_1+(y_1-y_2)};\frac{x_2u(\beta)+(x_1+y_1-y_2-x_2)u(\mathbf{0})}{x_2+(x_1+y_1-y_2-x_2)}\right)=\\
&=(x_1+y_1-y_2)u\left(\mathrm{Vp}_{x_1,y_1-y_2}(\alpha,\mathbf{1})\wedge\mathrm{Vp}_{x_2,x_1+y_1-y_2-x_2}(\beta,\mathbf{0})\right)+y_2
\end{aligned}
\tag{11.23}
$$

A pokud $x_1+y_1-y_2<x_2$:

$$
\max\left(x_1u(\alpha)+y_1-y_2;x_2u(\beta)\right)+y_2=x_2\max\left(\frac{x_1u(\alpha)+y_1-y_2}{x_2};u(\beta)\right)+y_2
\tag{11.24}
$$

$$
\min\left(x_1u(\alpha)+y_1-y_2;x_2u(\beta)\right)+y_2=x_2\min\left(\frac{x_1u(\alpha)+y_1-y_2}{x_2};u(\beta)\right)+y_2
\tag{11.25}
$$

Výrazy uvnitř maxim a minim jsou sice v rozmezí $\langle 0;1\rangle$, ale abychom je mohli převést na ohodnocení logických formulí, musíme opět zavést logické konstanty. Potom už snadno tyto výrazy převedeme na ohodnocení váženého průměru.

$$
\begin{aligned}
&x_2\max\left(\frac{x_1u(\alpha)+y_1-y_2}{x_2};u(\beta)\right)+y_2=\\
&=x_2\max\left(\frac{x_1u(\alpha)+(y_1-y_2)u(\mathbf{1})+(x_2-x_1-y_1+y_2)u(\mathbf{0})}{x_2};u(\beta)\right)+y_2=\\
&=x_2\max\left(\frac{x_1u(\alpha)+(y_1-y_2)u(\mathbf{1})+(x_2-x_1-y_1+y_2)u(\mathbf{0})}{x_1+(y_1-y_2)+(x_2-x_1-y_1+y_2)};u(\beta)\right)+y_2=\\
&=x_2u\left(\mathrm{Vp}_{x_1,y_1-y_2,x_2-x_1-y_1+y_2}(\alpha,\mathbf{1},\mathbf{0})\vee\beta\right)+y_2
\end{aligned}
\tag{11.26}
$$

$$
\begin{aligned}
& x_2 \min\left(\frac{x_1 u(\alpha) + y_1 - y_2}{x_2}; u(\beta)\right) + y_2 = \\
& = x_2 \min\left(\frac{x_1 u(\alpha) + (y_1 - y_2)u(\mathbf{1}) + (x_2 - x_1 - y_1 + y_2)u(\mathbf{0})}{x_2}; u(\beta)\right) + y_2 = \\
& = x_2 \min\left(\frac{x_1 u(\alpha) + (y_1 - y_2)u(\mathbf{1}) + (x_2 - x_1 - y_1 + y_2)u(\mathbf{0})}{x_1 + (y_1 - y_2) + (x_2 - x_1 - y_1 + y_2)}; u(\beta)\right) + y_2 = \\
& = x_2 u\left(\mathrm{Vp}_{x_1, y_1 - y_2, x_2 - x_1 - y_1 + y_2}(\alpha, \mathbf{1}, \mathbf{0}) \wedge \beta\right) + y_2
\end{aligned} \tag{11.27}
$$

A tedy každý prvek výstupu můžeme interpretovat jako fuzzy logickou kvaziformuli. Ještě můžeme provést následující úpravu:

$$xu(\alpha) + y = u(\alpha)(x + y) + (1 - u(\alpha))y \tag{11.28}$$

a zjistíme, že výsledek je konvexní kombinací dvou hodnot, mezi nimiž přecházíme na základě ohodnocení nějaké fuzzy logické formule.

Nic nám ale nezaručuje, že výsledná neuronová síť bude odpovídat nějaké jednoduché fuzzy logické formuli. Na druhou stranu, když máme nějaký odhad, jak zhruba by mohla výsledná formule vypadat, můžeme podle ní inicializovat hodnoty v neuronové síti. Tím se tu ale zabývat nebudeme, protože v této práci porovnáváme mezi sebou metody, které na začátku nevědí nic, nebo jenom velmi málo.

11.2 Vlastnosti neuronové sítě

U této neuronové sítě si můžeme všimnout několika zvláštních vlastností. Když si vezmeme první dvě vrstvy, můžeme z nich vyčíst informace o tom, jaké časové úseky jsou pro predikci časových řad důležité. Můžeme data z těchto dvou vrstev vzít a použít je v jiné síti, která předpovídá jinou časovou řadu, ve které platí podobné zákonitosti, což může aspoň zpočátku zlepšit predikční schopnosti neuronové sítě, která má zatím málo dat. Také je možné nahradit pouze poslední vrstvu.

V logistické regresi týkající se zobrazení času na parabolu si můžeme všimnout, že se tu detekují náhlé změny a výjimečné události. Pokud můžeme vyloučit možnost, že by tyto jevy byly periodické, a jsou daleko v minulosti, můžeme data před náhlou změnou nebo během výjimečné události vymazat, čímž se může predikční schopnost zlepšit.

Další výhodou je, že tuto neuronovou síť můžeme použít stejně pro předpověď skaláru i libovolně dlouhého vektoru, ať už spolu jednotlivé složky

souvisejí jakkoliv. Přitom neuronová síť si dokáže sama zjistit souvislost mezi jednotlivými složkami předpovídaného vektoru. Pokud se snažíme předpovídat hustotu chodců na nějaké chodbě, můžeme chodbu rozparcelovat na malá okna a hustoty v jednotlivých oken uspořádat do vektoru. Metody, které modelují hustotu chodců jako pravděpodobnostní rozdělení v prostoru a čase, vědí rovnou, která okna na sebe navazují, a tedy budou mít podobnou hustotu. Jenomže tyto metody si neporadí v situaci, kdy v prostředí jsou umístěny zdi, přičemž hustota chodců na každé straně je diametrálně odlišná.

Ještě horší je situace, kdy máme několikapodlažní budovu. Buď budeme modelovat každé podlaží zvlášť, anebo musíme hustotu modelovat pomocí nějakého rozdělení, které respektuje nespojitost mezi jednotlivými podlažími a správně modeluje schodiště. Pokud neuděláme ani jedno, můžeme dostat použitelnou předpověď, ale je otázka, co vlastně modelujeme. Například jaký má význam předpověď pro bod nacházející se mezi prvním a druhým podlažím? A aby toho nebylo málo, v některých budovách se nacházejí mezipatra, která jsou prostorově omezená, takže předpověď pro ně má význam pouze v určitém omezeném prostoru.

Tím ale podivné prostory nekončí. Uvažujme mimoúrovňovou křižovatku, na které předpovídáme hustotu vozidel. Dvourozměrný model nemůžeme použít, protože by se některé části vozovky promítly na sebe. Trojrozměrný model nemůžeme použít, protože smysl má pouze předpověď na vozovce, ale co se děje nad ní nebo pod ní, to není definováno. Můžeme si vozovku rozdělit na několik částí a každou z nich předpovídat zvlášť, ale není jasné, kde ji máme rozdělit, protože přechod mezi jednotlivými patry je pozvolný.

Pokud předpovídáme hustotu vozidel v nějaké dopravní síti, začneme mít problém s tím, že předpověď mimo dopravní linky nemá smysl. Pokud se jedná o síť silnic, můžeme potřebovat předpověď pro každý silniční pruh, takže model musí být dostatečně přesný, aby dokázal jednotlivé pruhy rozlišit. Potom se může ještě stát, že předpovídáme pouze vytíženost jednotlivých linek, a tedy nemáme informaci o tom, kde přesně se vozidla na linkách nacházejí. Tady už selhávají jakékoliv pokusy o zavedení euklidovského prostoru.

Také se může stát, že prostor máme už předem rozparcelovaný, ale ne tak, jak bychom chtěli nebo potřebovali. Může se stát, že známe pouze součty v jednotlivých správních celcích. Pokud chceme zavést nějaký prostor, tak si musíme data do jednotlivých správních celků někam umístit. To ale není triviální úloha, protože správní celky mají různé tvary a velikosti, nemusejí být jednoduše souvislé a ani nemusejí být souvislé. Můžeme udělat to, že každý naměřený údaj umístíme do daného správního celku náhodně s rovnoměrným rozdělením, ale potom budou všechny sousední správní celky ovlivňovány stejně, i když ve skutečnosti jsou všechny naměřené údaje koncentrovány

kolem jednoho bodu poblíž hranic.

Z předchozích příkladů vidíme, že zavádět prostor v predikci kvaziperiodických časových řad se vždy nevyplatí. Vytvářet model pro každou složku vektoru nezávisle se také nevyplácí, protože se jednotlivé složky nemění nezávisle na sobě, a tedy dostáváme spoustu podobných modelů. Navržená neuronová síť nabízí kompromis mezi těmito přístupy. Musí sice udělat nějakou práci navíc, aby zjistila, jaká prostorová okna spolu souvisejí a jakým způsobem, ale poradí si i ve složitějších prostorech.

Kapitola 12

Aplikace metod na úlohu s dveřmi

12.1 Metoda momentů (MM-vM a MM-kC)

Perioda byla nastavena na jeden týden a byly zkoušeny počty shluků 5 a 10, protože máme pět pracovních dnů v týdnu. Pro von Misesovo rozdělení solver v případě pěti shluků nenašel řešení pro soustavu týkající se rozdělení záporných dat a v případě deseti shluků nenašel řešení ani pro soustavu týkající se rozdělení kladných dat, a tak byla perioda změněna na jeden den a počty shluků byly nastaveny na 2, 4 a 6. Byla zvolena sudá čísla, aby bylo možné použít stejný počet reálných a imaginárních částí momentů. V metodě MM-kC solver našel řešení pouze pro denní periodu a počty shluků 1, 2 a 3.

12.2 EM algoritmus (EM-vM a EM-kC)

Perioda byla nastavena na jeden týden a byly zkoušeny počty shluků 5, 10, 15, 20, 25 a 30, protože máme pět pracovních dnů v týdnu a zvolené hodnoty jsou násobky pěti.

12.3 Adaboost$_{\mathbb{R}}$

Metoda Adaboost$_{\mathbb{R}}$ byla vyzkoušena pro 1, 2, 3 a 4 periodicity. Počet iterací byl nastaven na 20.

12.4 Neuronová síť (NS)

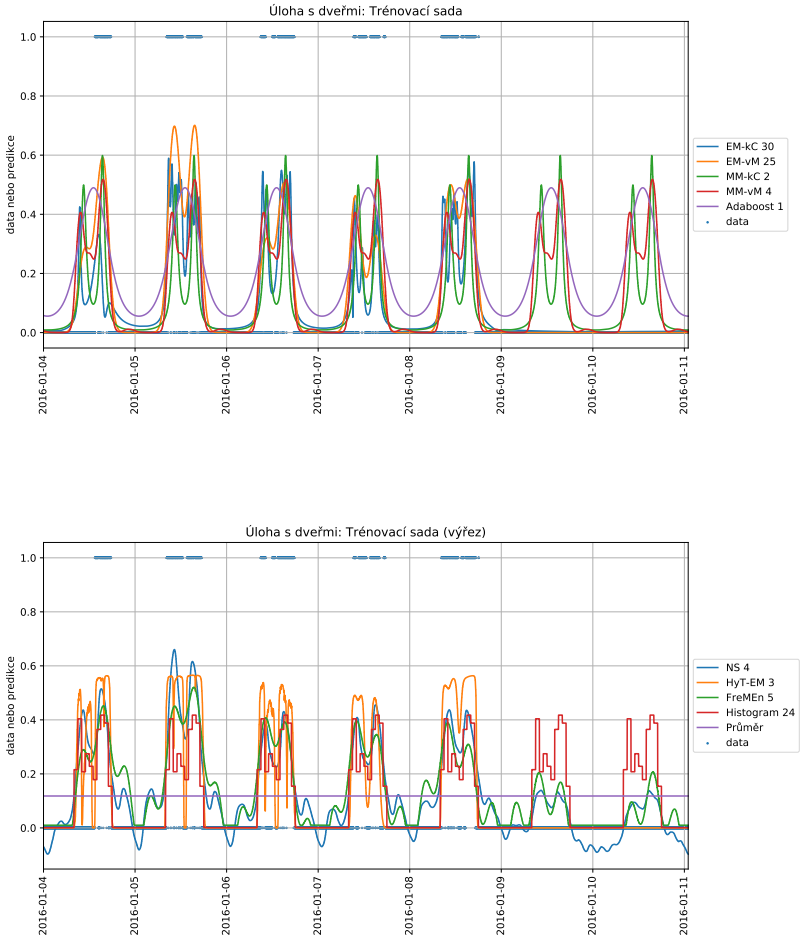
Použitá neuronová síť používá čtyři periodicity a zobrazení na parabolu (vzorec 11.2) bylo zakázáno. Logistická vrstva hledá na každé kružnici 7 logistických regresí a za ní následuje fuzzy logická vrstva. Potom následuje lineární vrstva, která má 12 výstupů, za ní fuzzy logická vrstva, za ní lineární vrstva s 5 výstupy, za ní fuzzy logická vrstva a poslední vrstva je lineární a má 1 výstup.

Síť byla trénována gradientní metodou s velikostí dávky 100 a s délkou kroku 0.01 do doby, než střední kvadratická chyba nebyla 30 epoch po sobě menší než 1.2. Potom se délka kroku snížila na 0.001. Doba trénování byla 10000 epoch.

12.5 Srovnání výsledků

Metoda FreMEn [8] byla vyzkoušena s jednou až pěti periodicitami, metoda HyT-EM [9] byla vyzkoušena s jedním až pěti shluky a histogram byl vyzkoušen s periodou jeden den a s počty košů 1, 4, 12 a 24. Výsledky jsou zobrazeny v grafech 12.1, 12.2, 12.3 a 12.4.

Nejlépe vychází předpověď pomocí EM algoritmu pro směs kruhových Cauchyových rozdělení. Zvláštní je, že EM algoritmus pro von Misesovo rozdělení vyšel o něco hůř, což by mohlo být podnětem pro další výzkum. Na rozdíl od metody HyT-EM se předpověď výrazně nezhoršuje s rostoucím počtem shluků. Zřejmě je to proto, že EM algoritmus používáme korektně. 15 shluků stačí metodě EM-kC, aby pokořila HyT-EM. Složitost obou modelů je srovnatelná, protože metoda HyT-EM hledá shluky v prostorech vyšších rozměrů a velikost kovarianční matice roste kvadraticky s dimenzí prostoru.

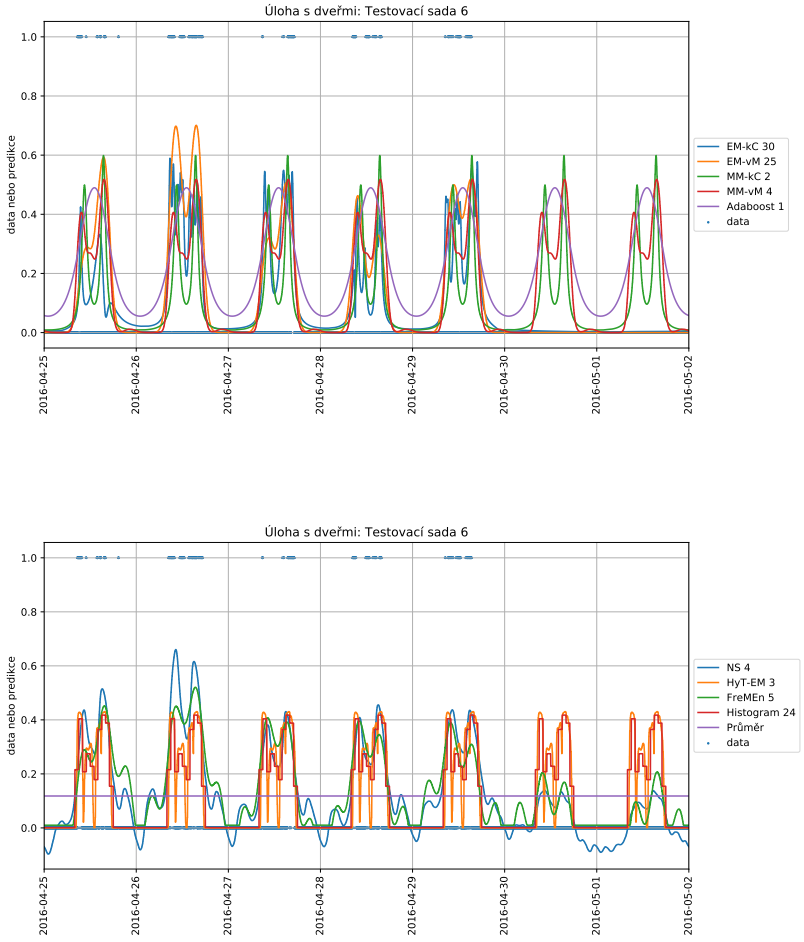


Obrázek 12.1: Úloha s dveřmi: predikce použitých metod pro trénovací sadu

Neuronová síť vyšla poměrně dobře, mohla by vyjít lépe, kdyby byla trénována delší dobu. Už takto ale trénování neuronové sítě zabralo mnohem déle než trénování ostatních metod. Doba potřebná k výpočtu předpovědi je také výrazně delší. I když neuronová síť získala svoje umístění za cenu velké složitosti a časové náročnosti, nesmíme zapomínat, že natrénovaná neuronová síť nám může usnadnit řešení dalších úloh, což ostatní modely neumožňují.

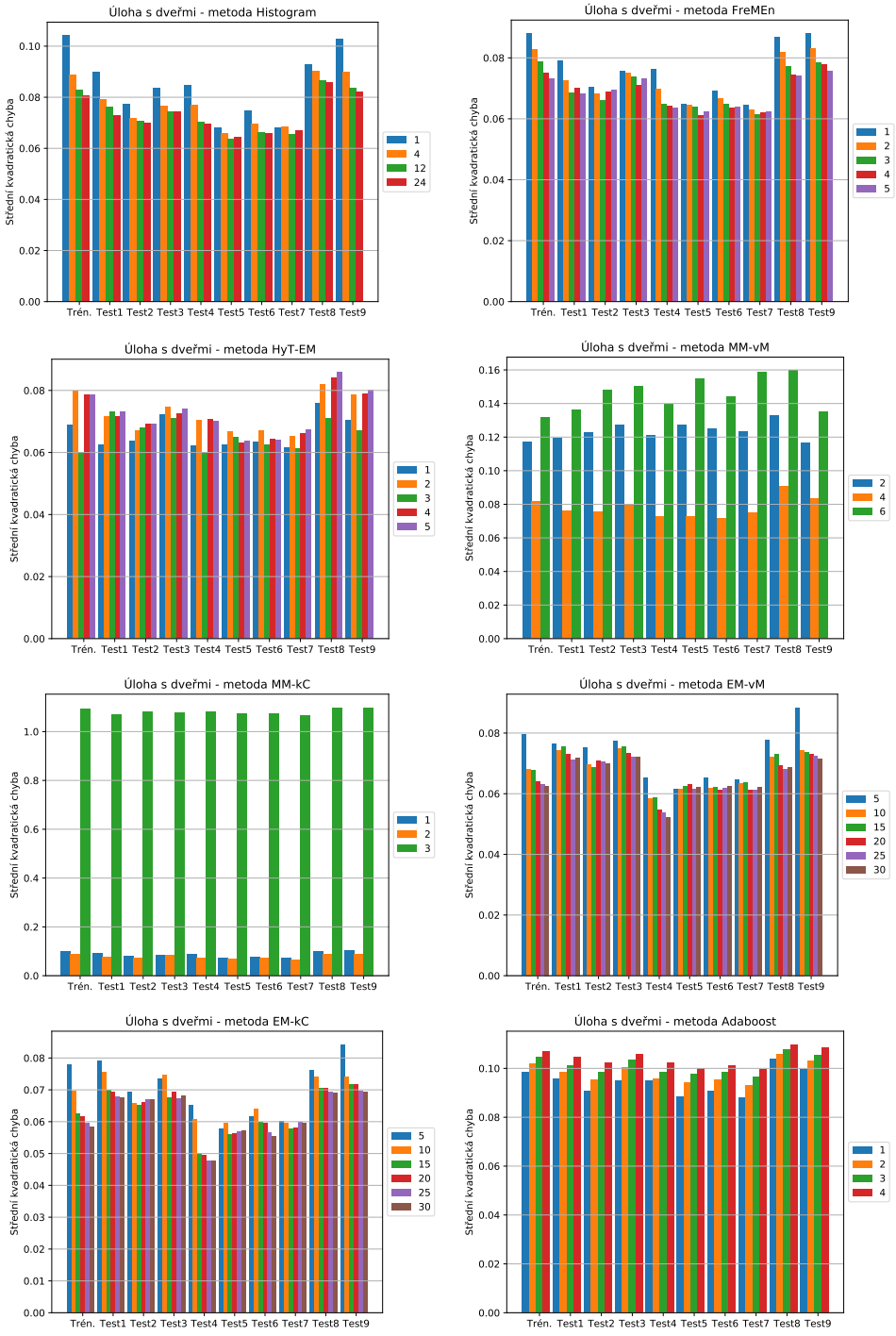
Metoda momentů vyšla špatně, protože řešení tak velké soustavy rovnic není zvládnuté zdaleka tak dobře jako postup u ostatních metod. Z výsledků metody EM-kC pro tři shluky je patrné, že solver sice našel řešení soustavy, ale neodpovídalo zadaným datům.

Překvapivě velká chyba vyšla u metody Adaboost$_{\mathbb{R}}$. Navíc se tato chyba

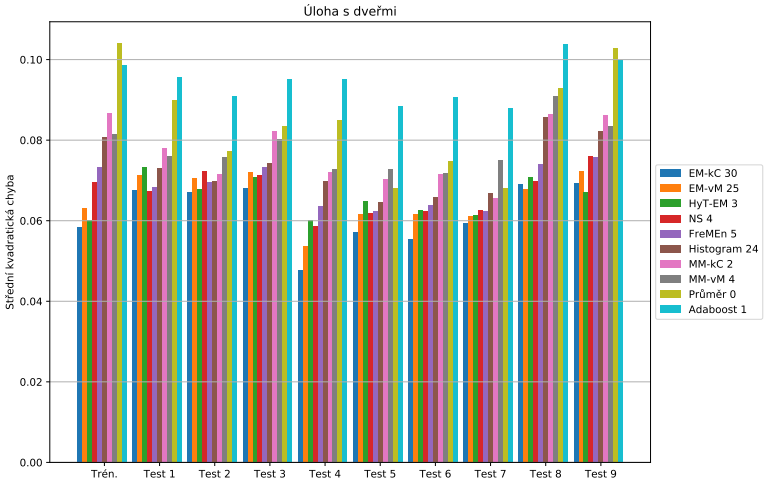


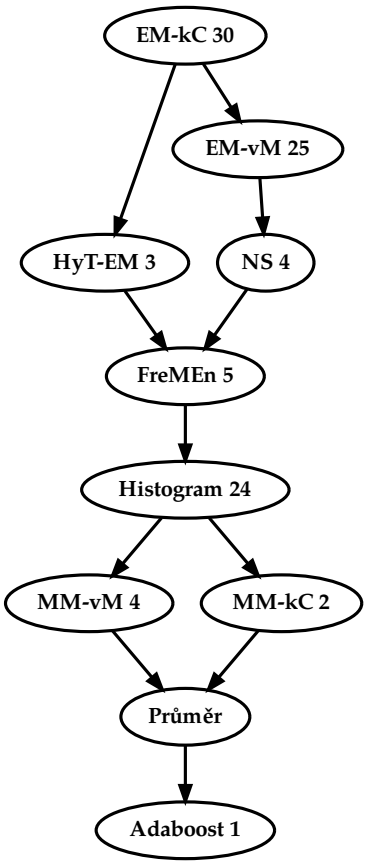
Obrázek 12.2: Úloha s dveřmi: predikce použitých metod pro testovací sadu 6

zhoršovala s rostoucím počtem použitých periodicit. V tabulkách 12.1 až 12.4 je uvedeno, jaké klasifikátory byly využity. Z hodnot je vidět, že v této úloze se Adaboost$_{\mathbb{R}}$ neblíží dobrému řešení, ale není zřejmé proč. To je překvapivý závěr a mohl by být podnětem pro další zkoumání.



Obrázek 12.3: Úloha s dveřmi: střední kvadratická chyba pro metody s parametry





Obrázek 12.4: Úloha s dveřmi: Šipka z metody A do metody B znamená, že metoda A je lepší než metoda B podle kritéria uvedeného v kapitole 2.4.

i	w_i	$T_i[s]$	u_i^T
1	0.77722	86400	(-2.64727; 0.833045; 3.1883)
2	0.347546	86400	(-2.0159; 0.587087; 2.0467)
3	0.231957	86400	(-1.71799; 0.484356; 1.66644)
4	0.174441	86400	(-1.53138; 0.423871; 1.45105)
5	0.139607	86400	(-1.39905; 0.382584; 1.30565)
6	0.116147	86400	(-1.29742; 0.35174; 1.19745)
7	0.0993229	86400	(-1.21619; 0.327563; 1.11349)
8	0.0866802	86400	(-1.14934; 0.307957; 1.04605)
9	0.0767844	86400	(-1.09309; 0.291654; 0.989246)
10	0.0688597	86400	(-1.04353; 0.277463; 0.940903)
11	0.0623402	86400	(-1.00045; 0.265231; 0.898371)
12	0.0569668	86400	(-0.963589; 0.254798; 0.862446)
13	0.0524	86400	(-0.929209; 0.245165; 0.830106)
14	0.0484507	86400	(-0.898456; 0.236599; 0.800203)
15	0.0450919	86400	(-0.871826; 0.229177; 0.774735)
16	0.0421281	86400	(-0.846698; 0.222226; 0.750899)
17	0.0395163	86400	(-0.822762; 0.215646; 0.729117)
18	0.0372023	86400	(-0.801459; 0.209795; 0.709201)
19	0.0351545	86400	(-0.78257; 0.204604; 0.691374)
20	0.033295	86400	(-0.764198; 0.199588; 0.67433)

Tabulka 12.1: Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro jednu periodicitu

i	w_i	$T_i[s]$	u_i^T
1	0.77722	86400	(-2.64727; -0.833045; -3.1883)
2	0.347546	86400	(-2.0159; -0.587087; -2.0467)
3	0.231957	86400	(-1.71799; -0.484356; -1.66644)
4	0.174441	86400	(-1.53138; -0.423871; -1.45105)
5	0.159427	604800	(0.950674; -0.552374; -0.637479)
6	0.130655	86400	(-1.42264; -0.392207; -1.2498)
7	0.110149	86400	(-1.32248; -0.361822; -1.15424)
8	0.114545	604800	(0.861411; -0.482794; -0.471328)
9	0.0921883	86400	(-1.25364; -0.343996; -1.05075)
10	0.0930362	604800	(0.791318; -0.435822; -0.404539)
11	0.0798446	86400	(-1.19499; -0.328985; -0.973504)
12	0.0784676	604800	(0.737729; -0.401375; -0.355408)
13	0.0706742	86400	(-1.14327; -0.315788; -0.912815)
14	0.0639372	86400	(-1.09207; -0.300736; -0.872132)
15	0.0674508	604800	(0.697748; -0.37547; -0.307042)
16	0.0579925	86400	(-1.05348; -0.291251; -0.828522)
17	0.0596883	604800	(0.661296; -0.353401; -0.28092)
18	0.0530751	86400	(-1.01711; -0.282237; -0.790767)
19	0.049119	86400	(-0.982643; -0.272094; -0.763618)
20	0.0534039	604800	(0.632162; -0.335553; -0.253633)

Tabulka 12.2: Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro dvě periodicity

i	w_i	$T_i[s]$	u_i^T
1	0.77722	86400	(-2.64727; -0.833045; -3.1883)
2	0.347546	86400	(-2.0159; -0.587087; -2.0467)
3	0.231957	86400	(-1.71799; -0.484356; -1.66644)
4	0.174441	86400	(-1.53138; -0.423871; -1.45105)
5	0.164809	21600	(-0.719941; -0.829062; -0.687033)
6	0.148969	604800	(0.961158; -0.554376; -0.565543)
7	0.127198	86400	(-1.47174; -0.406377; -1.22324)
8	0.107833	86400	(-1.36699; -0.374943; -1.1336)
9	0.115869	21600	(-0.642742; -0.757962; -0.47179)
10	0.109076	604800	(0.879372; -0.490639; -0.396042)
11	0.0921697	86400	(-1.32446; -0.365207; -1.03181)
12	0.0934968	21600	(-0.584511; -0.697428; -0.382214)
13	0.0820827	86400	(-1.27432; -0.350462; -0.973061)
14	0.0890314	604800	(0.817797; -0.448445; -0.309285)
15	0.0731053	86400	(-1.21519; -0.335937; -0.913429)
16	0.0803929	21600	(-0.548897; -0.661677; -0.314824)
17	0.07607	604800	(0.764084; -0.41521; -0.265232)
18	0.0670869	86400	(-1.1835; -0.329043; -0.867136)
19	0.070245	21600	(-0.513542; -0.623071; -0.276894)
20	0.0619005	86400	(-1.14746; -0.318552; -0.832865)

Tabulka 12.3: Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro tři periodicity

i	w_i	$T_i[s]$	u_i^T
1	0.77722	86400	(-2.64727; -0.833045; -3.1883)
2	0.347546	86400	(-2.0159; -0.587087; -2.0467)
3	0.231957	86400	(-1.71799; -0.484356; -1.66644)
4	0.174441	86400	(-1.53138; -0.423871; -1.45105)
5	0.164809	21600	(-0.719941; -0.829062; -0.687033)
6	0.148969	604800	(0.961158; -0.554376; -0.565543)
7	0.127198	86400	(-1.47174; -0.406377; -1.22324)
8	0.107833	86400	(-1.36699; -0.374943; -1.1336)
9	0.115869	21600	(-0.642742; -0.757962; -0.47179)
10	0.109076	604800	(0.879372; -0.490639; -0.396042)
11	0.111311	302400	(-0.871396; 0.351984; -0.374143)
12	0.0915959	86400	(-1.3372; -0.367056; -1.01502)
13	0.0937778	604800	(0.848848; -0.454009; -0.33295)
14	0.091958	21600	(-0.589924; -0.702908; -0.344969)
15	0.0819319	86400	(-1.29236; -0.356955; -0.948576)
16	0.0926316	302400	(-0.808022; 0.336918; -0.283864)
17	0.0828754	604800	(0.821994; -0.426322; -0.279361)
18	0.073523	86400	(-1.23972; -0.342632; -0.886618)
19	0.0805718	21600	(-0.558746; -0.669419; -0.28063)
20	0.0812319	302400	(-0.758534; 0.324561; -0.238569)

Tabulka 12.4: Adaboost$_{\mathbb{R}}$: Seznam slabých klasifikátorů pro čtyři periodicity

Kapitola 13

Aplikace metod na úlohu s vytížeností

13.1 Metoda momentů (MM-vM a MM-kC)

Počty hráčů byly rozděleny na 7 shluků. Pro týdenní periodu metoda momentů nedoběhla, protože nenašla řešení soustavy ani po vyčerpání 10^5 pokusů. Metoda MM-vM byla vyzkoušena pro denní periodu a počty shluků 2, 4 a 6. Byla zvolena sudá čísla, aby bylo možné použít stejný počet reálných a imaginárních částí momentů. Metoda MM-kC doběhla pouze pro denní periodu a 1 shluk.

13.2 EM algoritmus (EM-vM a EM-kC)

Počty hráčů byly rozděleny na 7 shluků. Perioda byla nastavena na jeden týden a byly zkoušeny počty shluků 5, 10, 15, 20, 25 a 30, protože týden má pět pracovních dnů a zvolené hodnoty jsou násobky pěti.

13.3 Adaboost$_{\mathbb{R}}$

Počty hráčů byly rozděleny na 3 shluky. Při rozdělení na větší počet shluků by se stalo, že data by byla lineárně separovatelná a logistická regrese by proto vytvořila ostré skoky. Metoda Adaboost$_{\mathbb{R}}$ byla vyzkoušena pro 1, 2, 3 a 4 periodicity. Počet iterací byl nastaven na 5.

13.4 Neuronová síť (NS)

Použitá neuronová síť používá čtyři periodicity a zobrazení na parabolu (vzorec 11.2) bylo povoleno. Logistická vrstva hledá na každé kružnici a parabole 7 logistických regresí a za ní následuje fuzzy logická vrstva. Potom následuje lineární vrstva, která má 15 výstupů, za ní fuzzy logická vrstva, za ní lineární vrstva s 5 výstupy, za ní fuzzy logická vrstva a poslední vrstva je lineární a má 1 výstup.

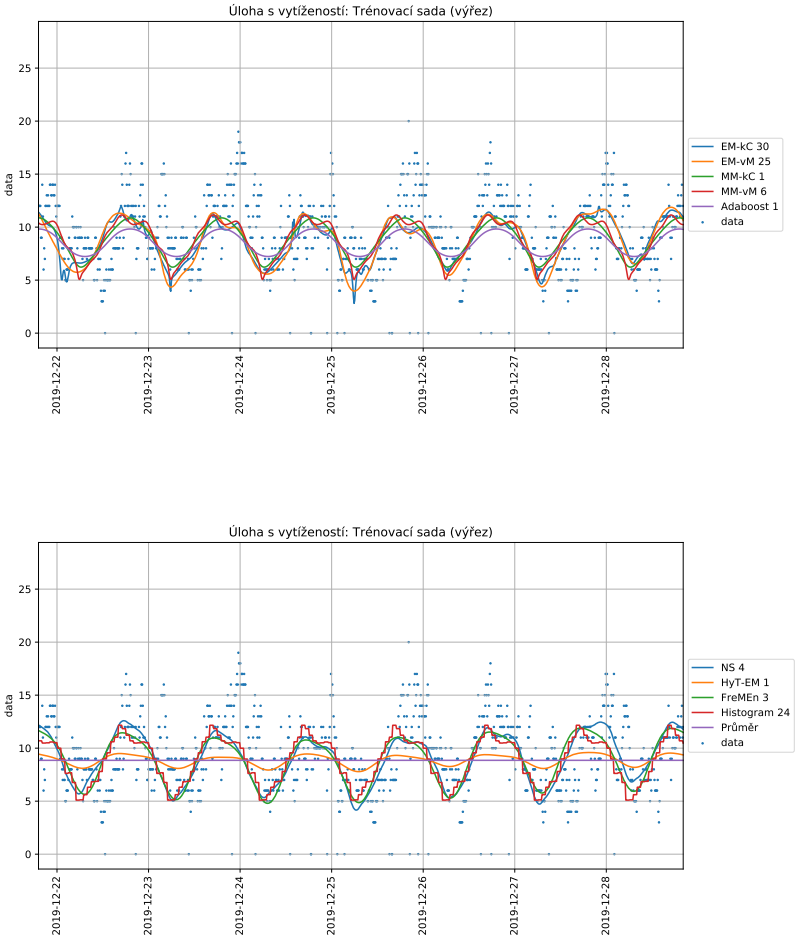
Síť byla trénována gradientní metodou s velikostí dávky 100 a s délkou kroku 0.001. Doba trénování byla 1000 epoch.

13.5 Srovnání výsledků

Metoda FreMEn [8] byla vyzkoušena s jednou až pěti periodicitami, metoda HyT-EM [9] byla vyzkoušena s jedním až pěti shluky a histogram byl vyzkoušen s periodou jeden den a počty košů 1, 4, 12 a 24. Výsledky jsou zobrazeny v grafech 13.1, 13.2, 13.3 a 13.4.

Metoda Adaboost$_{\mathbb{R}}$, která byla v předchozí úloze nejhorší, je v této úloze skoro nejlepší. Podobně jako v úloze s dveřmi, denní periodicita převládla nad ostatními a všechny slabé klasifikátory využily denní periodu, přestože metoda měla možnost si vybrat i jiné periodicity. To je důvod, proč se chyba neměnila v závislosti na složitosti metody.

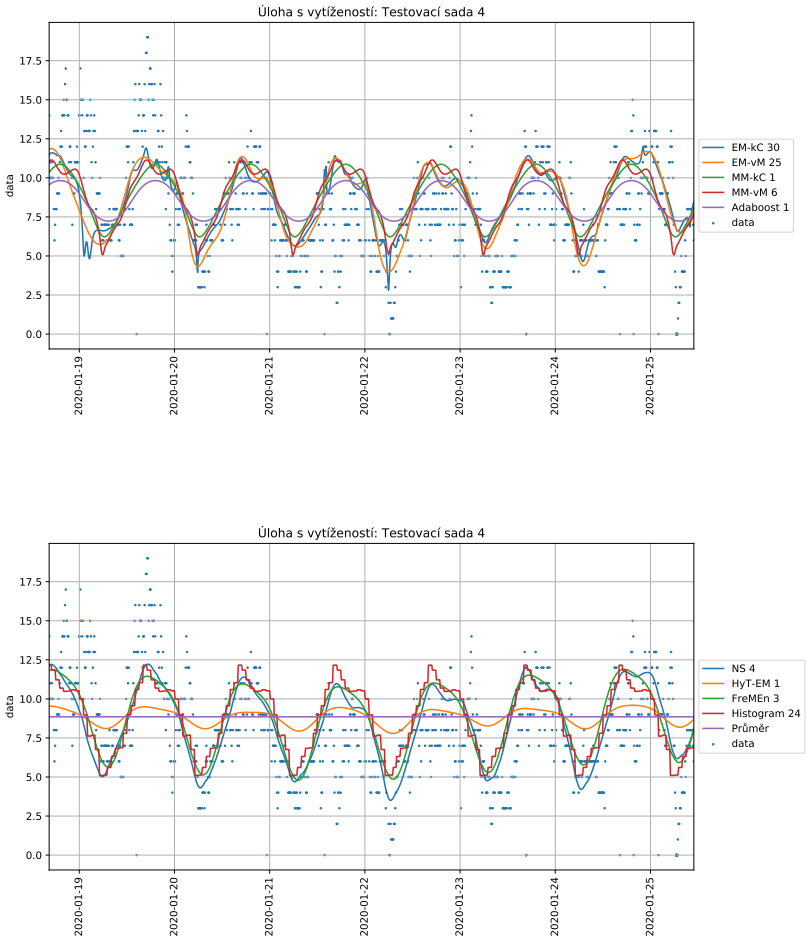
Také vyšel dobře EM algoritmus, ale tentokrát vyšel výrazně lépe pro směs von Misesových rozdělení. Neuronová síť vyšla také dobře, mohla by vyjít



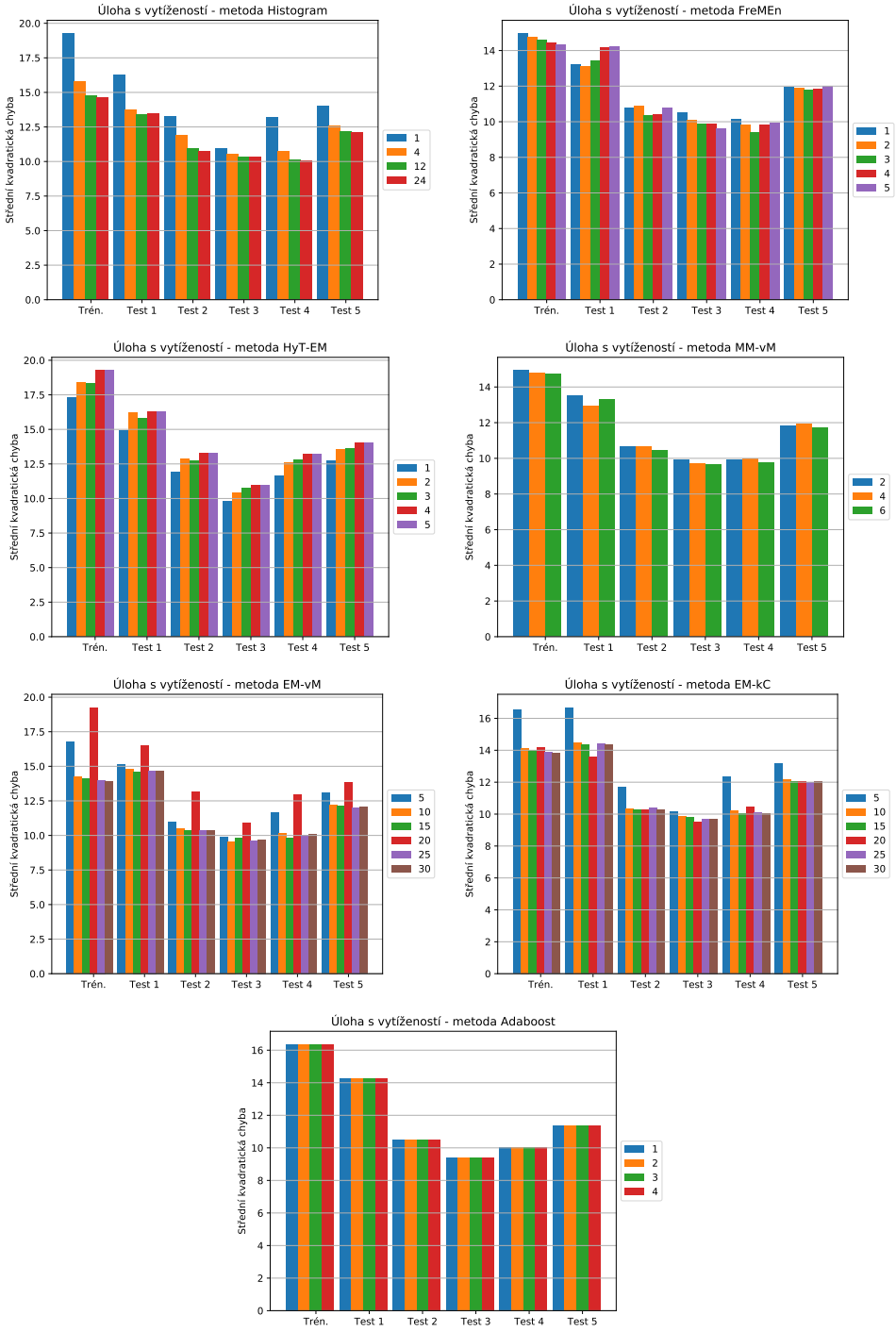
Obrázek 13.1: Úloha s vytížeností: predikce použitých metod pro trénovací sadu. Predikce metod EM-vM a MM-kc téměř splývají.

lépe, kdyby byla déle trénována. Už takto ale její trénování zabralo výrazně déle než trénování ostatních metod. Doba potřebná k výpočtu předpovědi je také výrazně delší. Opět neuronová síť získala svoje umístění za cenu velké složitosti a časové náročnosti, ale již natrénovaná neuronová síť může usnadnit řešení dalších úloh, což ostatní modely neumožňují.

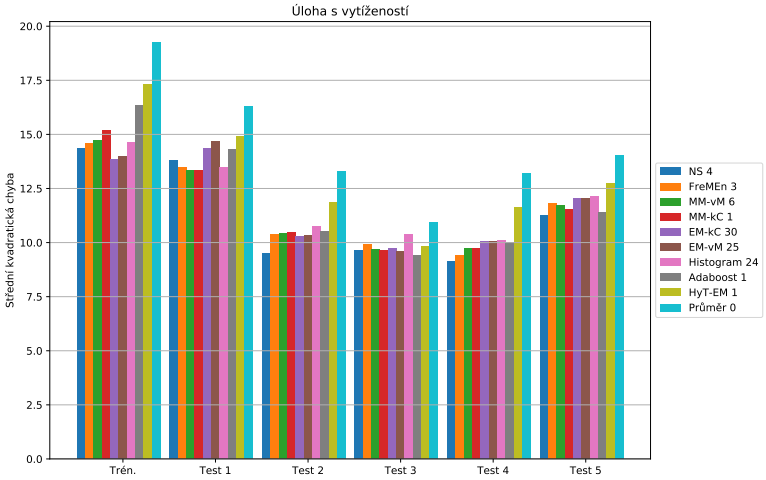
Překvapivé je, že předpovědi pomocí metody momentů vyšly výrazně lépe než pomocí EM-algoritmu, přestože jejich složitost je výrazně menší. A velmi dobře dopadla metoda MM-kC, přestože si musela vystačit jen s jedním shlukem.

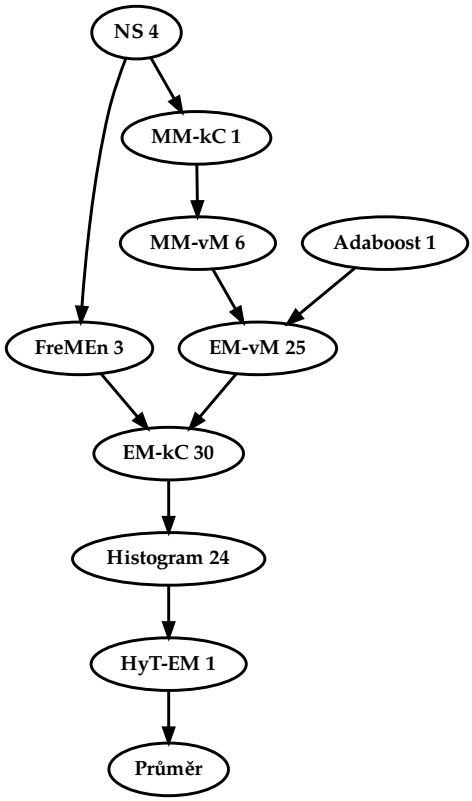


Obrázek 13.2: Úloha s vytížeností: predikce použitých metod pro testovací sadu 4. Predikce metod EM-vM a MM-kc téměř splývají.



Obrázek 13.3: Úloha s vytížeností: střední kvadratická chyba pro metody s parametry





Obrázek 13.4: Úloha s vytížeností: Šipka z metody A do metody B znamená, že metoda A je lepší než metoda B podle kritéria uvedeného v kapitole 2.4.

Kapitola 14

Aplikace metod na úlohu s chodci

14.1 Metoda momentů (MM-vM a MM-kC)

Zde uvažujeme prostor, který má jeden rozměr komplexní a dva rozměry reálné. Jedinou možností, jak provést metodu momentů, je rozdělit komplexní rozměr na dva reálné. To provedeno nebylo, protože výpočet by byl příliš složitý a metoda momentů dávala špatné výsledky už na předchozích dvou úlohách, které jsou výrazně jednodušší.

14.2 EM algoritmus (EM-vM a EM-kC)

Protože zde hraje roli i prostor, každý shluk má dvě nezávislé složky, přičemž ta, která modeluje prostor, má Gaussovo rozdělení. Perioda byla nastavena na jeden týden a byly zkoušeny počty shluků od jednoho do pěti.

14.3 Adaboost$_{\mathbb{R}}$

Adaboost$_{\mathbb{R}}$ je klasifikační metoda, která rozeznává od sebe pozitivní a negativní data. V této úloze máme pouze pozitivní data, takže Adaboost$_{\mathbb{R}}$ nelze uplatnit.

14.4 Neuronová síť (NS)

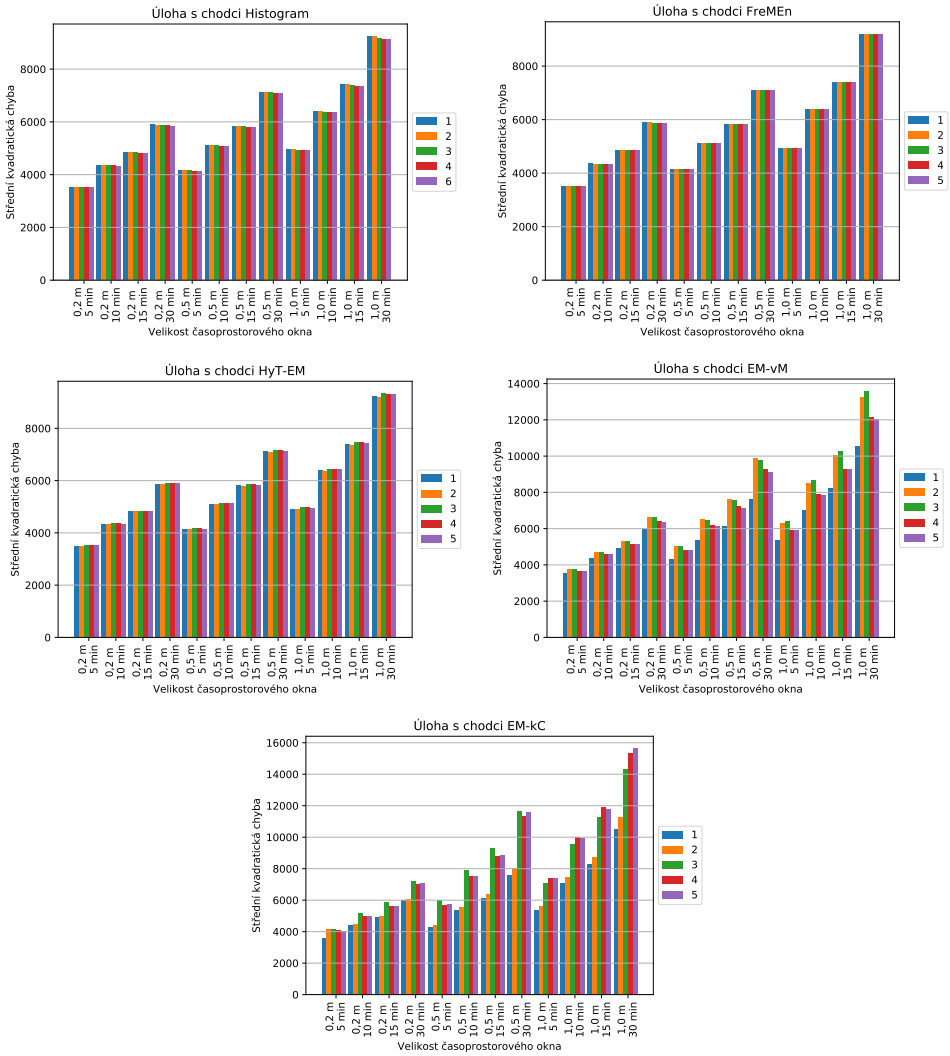
Zde byla vzata již natrénovaná neuronová síť z úlohy s dveřmi a její poslední lineární vrstva byla nahrazena lineární vrstvou, která má požadovaný počet výstupů. Dále byla trénována pouze poslední vrstva, která byla trénována gradientní metodou s velikostí dávky 30 a s délkou kroku 0.1. Doba trénování byla 200 epoch.

14.5 Srovnání výsledků

Metoda FreMEn [8] byla vyzkoušena s jednou až pěti periodicitami, metoda HyT-EM [9] byla vyzkoušena s jedním až pěti shluky a histogram byl vyzkoušen s periodou jeden den a počty košů 1, 2, 3, 4 a 6. Výsledky jsou zobrazeny v grafech 14.1 a 14.2.

Chyby jednotlivých metod se výrazně neliší, protože předpovídaná data mají velký šum. Ten je zřejmě způsoben tím, že počet chodců je celé číslo a celkově se na chodbě nevyskytuje dostatek chodců na to, aby jejich počet mohl být vnímán spojitě. I tak je ale možné předpověď zlepšit tím, že v časech, kdy se v daném prostoru vyskytne chodec častěji, je předpovězen o něco větší počet chodců.

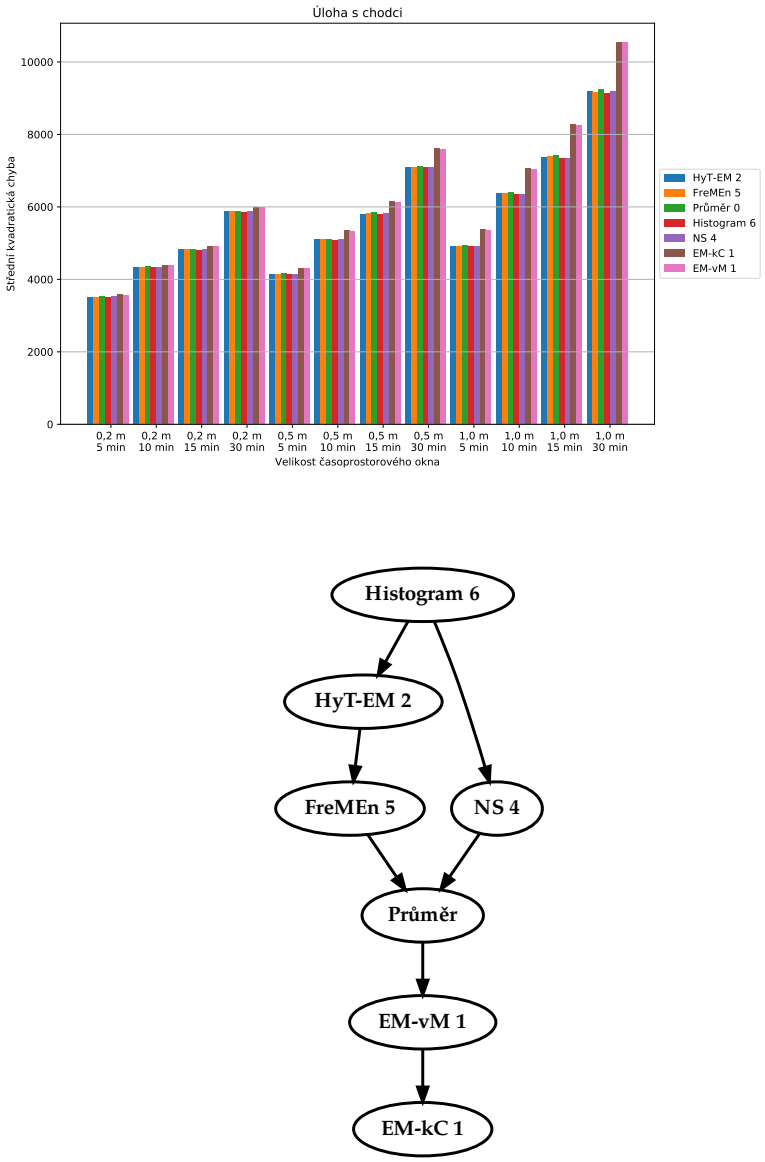
Překvapivé je, že odhad pomocí EM algoritmu vyšel špatně, dokonce ještě výrazně hůř než předpověď průměrem. Také je zvláštní, že chyba je nejmenší pro jeden shluk a v případě von Misesova rozdělení roste se zvyšujícím počtem shluků a v případě kruhového Cauchyova rozdělení nejdříve roste a potom klesá. I když tu na rozdíl od metody HyT-EM pracujeme s EM algoritmem korektně, předpověď je mnohem horší. Nalezené shluky nevypadají, jako by



Obrázek 14.1: Úloha s chodci: střední kvadratická chyba pro metody s parametry

vznikly chybou, jsou vypsány v tabulkách 14.1 a 14.2. Neúspěch je možná způsoben tím, že hledáme maximálně věrohodný model, ale metody srovnáváme podle střední kvadratické chyby. Naopak histogram i průměr optimalizují střední kvadratickou chybu v rámci možností omezených modelem.

Další překvapení je, že nejlepší metodou pro řešení této úlohy se jeví histogram, což je hned po průměru ta nejjednodušší metoda. Neuronová síť by mohla vyjít trochu lépe, kdyby byla déle a lépe trénována. V této úloze je neuronová síť svým způsobem postavena mimo soutěž, protože vychází z dat, které nasbírala v jiné úloze. Nicméně to ilustruje jednu její zajímavou vlastnost, a sice že je možné tuto síť přenášet mezi úlohami a šetřit tím čas při trénování. Proto také doba jejího trénování byla srovnatelná s dobou trénování ostatních metod.



Obrázek 14.2: Úloha s chodci: Šipka z metody A do metody B znamená, že metoda A je lepší než metoda B podle kritéria uvedeného v kapitole 2.4.

model	$EX[m]$	$EY[m]$	$DX[m^2]$	$DY[m^2]$
	$\mathrm{cov}(X,Y)[m^2]$	μ	κ	váha
EM-vM 1	1.98413	-0.712275	8.39444	4.1989
	-0.30417	-2.30132	3.25816	1
EM-vM 2	1.5719	-1.67123	1.31869	0.126617
	0.0259052	-1.91875	119.787	0.459375
	2.3344	0.102556	14.1397	6.21383
	-1.20595	-2.70211	2.99509	0.540625
EM-vM 3	1.56806	-1.7001	1.3572	0.118996
	0.0387416	-1.91865	129.167	0.441157
	1.10835	2.48571	0.254819	11.7213
	-0.482304	-2.59519	2.95002	0.161347
	2.8014	-0.914041	18.4164	0.40549
	0.235414	-2.70587	2.97268	0.397495
EM-vM 4	3.96556	-0.914912	10.6945	0.389633
	0.287212	-2.80684	3.23721	0.349304
	0.768762	-1.57619	0.467907	0.698283
	0.408327	-1.97131	300	0.19509
	2.0716	-1.65689	2.38732	0.095855
	-0.262367	-1.88329	118.217	0.280222
	-0.75004	2.16157	7.05778	11.9605
	5.25382	-2.56537	2.88413	0.175383
EM-vM 5	2.08287	-1.66545	2.42929	0.09269
	-0.255285	-1.88252	126.424	0.271776
	1.03877	3.19375	0.216702	11.4315
	-0.32698	-2.6664	2.95297	0.13199
	0.765085	-1.6052	0.500597	0.687361
	0.429058	-1.9722	300	0.19039
	0.272042	-0.910568	16.2961	0.397909
	-0.0379645	-2.58045	3.08622	0.201991
	5.29956	-0.940121	7.00131	0.408032
	0.605362	-2.85588	3.15561	0.203854

Tabulka 14.1: EM-vM: seznam nalezených shluků

model	$EX[m]$	$EY[m]$	$DX[m^2]$	$DY[m^2]$
	$\mathrm{cov}(X,Y)[m^2]$	$\Im\mu$	$\Re\mu$	váha
EM-kC 1	1.98413	-0.712275	8.39444	4.1989
	-0.30417	-0.313428	-0.761118	1
EM-kC 2	1.61354	-1.64084	1.40065	0.147085
	0.0275266	-0.336544	-0.872976	0.515698
	2.37875	0.276492	15.5397	6.61759
	-1.41399	-0.58229	-0.2153	0.484302
EM-kC 3	1.60569	-1.69478	1.3719	0.117633
	0.0371593	-0.338572	-0.87832	0.46977
	1.08559	2.66223	0.238009	11.5367
	-0.413294	-0.491277	-0.310254	0.15423
	2.82552	-0.86892	19.3455	0.390884
	0.226473	-0.581819	-0.223814	0.376
EM-kC 4	2.11751	-1.66119	2.50053	0.0894742
	-0.252454	-0.291475	-0.891338	0.307223
	1.05144	3.09549	0.228929	11.6123
	-0.379345	-0.593663	-0.186339	0.134984
	2.9124	-0.853652	20.1222	0.3782
	0.272124	-0.599107	-0.206794	0.353069
	0.798042	-1.55508	0.492816	0.634294
	0.390628	-0.388008	-0.888039	0.204724
EM-kC 5	2.10404	-1.67069	2.49626	0.0875619
	-0.242028	-0.291951	-0.8959	0.28946
	1.02916	3.51643	0.21132	10.9948
	-0.31503	-0.574426	-0.225626	0.121769
	3.30997	-0.951189	15.4253	0.37728
	0.503769	-0.783546	0.0128305	0.207298
	0.772316	-1.59962	0.501471	0.66965
	0.419297	-0.389649	-0.889839	0.191714
	2.18994	-0.806385	20.8703	0.40808
	0.0134221	-0.406095	-0.608407	0.189759

Tabulka 14.2: EM-kC: seznam nalezených shluků

Kapitola 15

Závěr

Byly prozkoumány nové statistické metody předpovídání kvaziperiodických časových řad, některé jsou úspěšné, jiné moc úspěšné nejsou. Není však vyloučeno, že by nešly zdokonalit. Podle srovnávací metody uvedené v kapitole 2.4 jsou některé metody v některých úlohách lepší než metody FreMEn [8] a HyT-EM [9]. Překvapivě chyba některých metod na některých datech nečekaně roste se složitostí metody. Není zřejmé, proč k takovým jevům dochází, a je to zajímavý podnět pro další zkoumání.

15.1 Metoda momentů

Metoda momentů dopadla špatně v tom smyslu, že často nenašla žádné řešení. Ale když už nějaké řešení našla, tak někdy nebylo úplně k zahození. Kdyby se podařilo nalézt způsob, jak tyto rovnice řešit, který by zaručoval úspěšné řešení, znamenalo by to výrazné zlepšení. Výhodou této metody je, že odhad momentů lze uskutečnit už během sběru dat a doba řešení rovnic nezáleží na množství nasbíraných dat.

15.2 EM algoritmus

EM algoritmus dopadl dobře v úloze s dveřmi a úloze s vytížeností, ale nedopadl dobře v úloze s chodci. Tato metoda je v jistém smyslu podobná metodě HyT-EM [9], hlavní rozdíl je v tom, že se nesnažíme rozdělení na kružnici odhadovat pomocí směsi vícerozměrných Gaussových rozdělení, ale pomocí rozdělení na kružnici, takže je korektnější. Důvod, proč dopadl EM algoritmus tak špatně v úloze s chodci, zřejmě spočívá v tom, že optimalizuje věrohodnost a ne kvadratickou chybu.

15.3 Adaboost$_{\mathbb{R}}$

V úloze s dveřmi dopadl Adaboost$_{\mathbb{R}}$ nejhůře, zatímco v úloze s vytížeností dopadl skoro nejlépe. Plyne z toho, že Adaboost$_{\mathbb{R}}$ se hodí na předpověď spojitých řad, ale nehodí se na předpověď binárních řad. Adaboost$_{\mathbb{R}}$ umí jednotlivé slabé klasifikátory pouze sčítat s různými vahami, ale neumí provádět operace jako konjunkce nebo disjunkce. V případě úlohy s dveřmi by potřeboval nějak namodelovat, že dveře jsou otevřené, když je den a zároveň není víkend. Kdyby v pracovní dny v noci a o víkendu ve dne byly dveře trochu otevřené, už by to šlo namodelovat pomocí váženého součtu dvou slabých klasifikátorů, přičemž jeden by určoval, zda je den, a druhý by určoval, zda není víkend.

15.4 Neuronové sítě

Neuronová síť skončila ve všech na celkem dobrém umístění, ale nebyla úplně nejlepší. Také je nutno poznamenat, že ze všech metod skladuje nejvíce dat. Na rozdíl od ostatních metod má zajímavé vlastnosti, které ostatní metody nemají. Těmto vlastnostem je věnována samostatná podkapitola 11.2. Za zmínku zde stojí, že neuronovou síť je možné s drobnými úpravami přenášet z jedné úlohy do druhé, i když je výstup diametrálně odlišný. To může usnadnit situaci ve chvíli, kdy je od robota požadováno, aby fungoval v lidském prostředí, aniž by měl k dispozici delší řadu měření z daného prostředí.

Literatura

[1] Abramowitz, M., and Stegun, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (1964)*, ninth Dover printing, tenth GPO printing ed. Dover, New York, 1964.

[2] Biber, P., and Duckett, T. Experimental Analysis of Sample-Based Maps for Long-Term SLAM. *I. J. Robotic Res. 10.1177/0278364908096286 28* (01 2009), 20–33.

[3] Bustince, H., Castillo, A., Rodrigues, I., Fumanal, J., Fernandez, J., and Minárová, M. Effect of Different Distances in the Gravitational Clustering Algorithm. *Fifteenth International Conference on Fuzzy Set Theory and Applications FSTA* in Liptovský Ján (01 2020).

[4] Churchill, W., and Newman, P. Experience-based Navigation for Long-term Localisation. *The International Journal of Robotics Research 10.1177/0278364913499193 32*, 14 (2013), 1645–1661.

[5] Galassi, M. e. a. GNU Scientific Library Reference Manual `https://www.gnu.org/software/gsl/doc/html/index.html`, 2018.

[6] Hawes, N., Burbridge, C., Jovan, F., Kunze, L., Lacerda, B., Mudrová, L., Young, J., Wyatt, J., Hebesberger, D., Koertner, T., Ambrus, R., Bore, N., Folkesson, J., Jensfelt, P., Beyer, L., Hermans, A., Leibe, B., Aldoma, A., F., T., and Hanheide, M. The STRANDS Project: Long-Term Autonomy in Everyday Environments. *IEEE Robotics & Automation Magazine PP99 10.1109/MRA.2016.2636359* (04 2016).

[7] Kent, J. T., and Tyler, D. E. Maximum Likelihood Estimation for the Wrapped Cauchy distribution. *Journal of Applied Statistics 15*, 2 (1988), 247–254.

[8] Krajník, T., Fentanes, J. P., ao Santos, J., Kusumam, K., and Duckett, T. FreMEn: Frequency Map Enhancement for Long-Term Mobile Robot Autonomy in Changing Environments (2015). In review.

[9] Krajník, T., Vintr, T., Molina, S., Fentanes, J. P., Cielniak, G., Mozos, O., and Duckett, T. Wrapped Hypertime Representations for Long-term Autonomy of Mobile Robots. *IEEE Robotics and Automation Letters* (2019). To appear.

[10] Kucner, T., Saarinen, J., Lilienthal, A., and Magnusson, M. Conditional Transition Maps: Learning Motion Patterns in Dynamic Environments. *IEEE International Conference on Intelligent Robots and Systems in Tokyo 10.1109/IROS.2013.6696502* (11 2013).

[11] Mardia, K. V. K. V. *Directional Statistics*. Wiley Series in Probability and Statistics. Wiley, Chichester, (1999), 36-45.

[12] McCullagh, P. Conditional Inference and Cauchy Models. *Biometrika 79*, 2 (06 1992), 247–259.

[13] Nock, R., and Nielsen, F. A Real Generalization of Discrete AdaBoost. *Artif. Intell. 171* (2006), 25–41.

[14] Schou, G. Estimation of the Concentration Parameter in von Mises-Fisher Distributions. *Biometrika 65*, 2 (1978), 369–377.

[15] Sra, S. Directional Statistics in Machine Learning: A Brief Review. *arXiv:1605.00316* (2016), 2.

[16] Tipaldi, G. D., Meyer-Delius, D., and Burgard, W. Lifelong Localization in Changing Environments. *International Journal of Robotics Research 10.1177/0278364913502830 32* (12 2013), 1662–1678.

[17] Yan, Z., Duckett, T., and Bellotto, N. Online Learning for Human Classification in 3D LiDAR-based Tracking. *10.1109/IROS.2017.8202247* (09 2017), 864–871.

Příloha A

Důkaz některých vlastností kruhového Cauchyova rozdělení

Tyto důkazy jsem vymyslel nezávisle na [12].

Lemma 1: Pro $x \in \langle 0;1)$ a $t \in (-\pi;\pi)$ platí:

$$\int \frac{x^2-1}{2(x^2+1-2x\cos t)}\mathrm{d}t = \operatorname{arctg}\left(\frac{x+1}{x-1}\operatorname{tg}\frac{t}{2}\right)+C \qquad (A.1)$$

Důkaz: Nejdříve ukážeme, že výraz pod zlomkovou čarou není nulový.

$$2(x^2+1-2x\cos t) \geq 2(x^2+1-2x) = 2(x-1)^2 > 0 \qquad (A.2)$$

A nyní samotný důkaz.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\operatorname{arctg}\left(\frac{x+1}{x-1}\operatorname{tg}\frac{t}{2}\right)+C\right) = \frac{\frac{x+1}{x-1}\cdot\frac{1}{2\cos^2\frac{t}{2}}}{1+\frac{(x+1)^2}{(x-1)^2}\operatorname{tg}^2\frac{t}{2}} = \frac{\frac{x+1}{x-1}\cdot\frac{1}{2\cos^2\frac{t}{2}}}{1+\frac{(x+1)^2\sin^2\frac{t}{2}}{(x-1)^2\cos^2\frac{t}{2}}} =$$

$$= \frac{\frac{x+1}{x-1}\cdot\frac{1}{2\cos^2\frac{t}{2}}}{\frac{(x-1)^2\cos^2\frac{t}{2}+(x+1)^2\sin^2\frac{t}{2}}{(x-1)^2\cos^2\frac{t}{2}}} = \frac{(x+1)(x-1)\frac{1}{2\cos^2\frac{t}{2}}}{\frac{(x-1)^2\cos^2\frac{t}{2}+(x+1)^2\sin^2\frac{t}{2}}{\cos^2\frac{t}{2}}} =$$

$$= \frac{(x+1)(x-1)}{2\left((x-1)^2\cos^2\frac{t}{2}+(x+1)^2\sin^2\frac{t}{2}\right)} =$$

$$= \frac{x^2-1}{2\left(x^2\cos^2\frac{t}{2}+x^2\sin^2\frac{t}{2}+\cos^2\frac{t}{2}+\sin^2\frac{t}{2}+2x\sin^2\frac{t}{2}-2x\cos^2\frac{t}{2}\right)} =$$

$$= \frac{x^2-1}{2\left(x^2+1+2x\sin^2\frac{t}{2}-2x\cos^2\frac{t}{2}\right)} = \frac{x^2-1}{2(x^2+1-2x\cos t)}$$

Lemma 2: Pro všechna $n \in \mathbb{N}$ a $x \in \langle 0; 1)$ platí:

$$(x^2+1-2x\cos t)(-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos it) =$$
$$=(x^{2n+2}+1)\cos nt-(x^{2n}+1)x\cos(n+1)t \tag{A.3}$$

Důkaz: (indukcí podle n)
Základní krok: $n=1$

$$(x^2+1-2x\cos t)((x^2+1)\cos t+x) =$$
$$=x^3+x-2x^2\cos t+x^2(x^2+1)\cos t+(x^2+1)\cos t-2x(x^2+1)\cos^2 t =$$
$$=x^3+x-2x^2\cos t+x^4\cos t+x^2\cos t+x^2\cos t+\cos t-2x^3\cos^2 t-2x\cos^2 t =$$
$$=x^3+x+x^4\cos t+\cos t-2x^3\cos^2 t-2x\cos^2 t =$$
$$=x^3+x+x^4\cos t+\cos t-x^3(1+\cos 2t)-x(1+\cos 2t) =$$
$$=x^3+x+x^4\cos t+\cos t-x^3-x^3\cos 2t-x-x\cos 2t =$$
$$=x^4\cos t+\cos t-x^3\cos 2t-x\cos 2t =$$
$$=(x^4+1)\cos t-(x^2+1)x\cos 2t \tag{A.4}$$

Indukční krok: Předpokládejme, že pro nějaké $n \in \mathbb{N}$ Lemma 2 platí. Potom:

$$(x^2+1-2x\cos t)(-x^{n+1}+\sum_{i=0}^{n+1}(x^{2i}+1)x^{n-i+1}\cos it) =$$
$$=(x^2+1-2x\cos t)\big((x^{2n+2}+1)\cos(n+1)t+x(-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos it)\big) \overset{\text{I.P.}}{=}$$
$$\overset{\text{I.P.}}{=}(x^2+1-2x\cos t)(x^{2n+2}+1)\cos(n+1)t+$$
$$+x\big((x^{2n+2}+1)\cos nt-(x^{2n}+1)x\cos(n+1)t\big) =$$
$$=(x^{2n+2}+1)x^2\cos(n+1)t+(x^{2n+2}+1)\cos(n+1)t-(x^{2n+2}+1)x\cos nt-$$
$$-(x^{2n+2}+1)x\cos(n+2)t+(x^{2n+2}+1)x\cos nt-(x^{2n}+1)x^2\cos(n+1)t =$$
$$=(x^{2n+2}+1)x^2\cos(n+1)t+(x^{2n+2}+1)\cos(n+1)t-$$
$$-(x^{2n+2}+1)x\cos(n+2)t-(x^{2n}+1)x^2\cos(n+1)t =$$
$$=x^{2n+4}\cos(n+1)t+x^2\cos(n+1)t+x^{2n+2}\cos(n+1)t+\cos(n+1)t-$$
$$-x^{2n+3}\cos(n+2)t-x\cos(n+2)t-x^{2n+2}\cos(n+1)t-x^2\cos(n+1)t =$$
$$=x^{2n+4}\cos(n+1)t+\cos(n+1)t-x^{2n+3}\cos(n+2)t-x\cos(n+2)t =$$
$$=(x^{2n+4}+1)\cos(n+1)t-(x^{2n+2}+1)x\cos(n+2)t \tag{A.5}$$

A tedy lemma platí.

Lemma 3: Pro všechna $n \in \mathbb{N}$ a $x \in \langle 0; 1), t \in \mathbb{R}$ platí:

$$\frac{x^{n+1}(x^2-1)}{x^{2n+2}+1}\frac{\cos(n+1)t}{x^2+1-2x\cos t}-\frac{x^n(x^2-1)}{x^{2n}+1}\frac{\cos nt}{x^2+1-2x\cos t} =$$
$$=-\frac{x^n(x^2-1)}{(x^{2n}+1)(x^{2n+2}+1)}(-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt) \tag{A.6}$$

Důkaz: Nejdříve ukážeme, že nedochází k dělení nulou.

$$x^2 + 1 - 2x\cos t \geq x^2 + 1 - 2x = (x-1)^2 > 0 \tag{A.7}$$

A teď důkaz samotného tvrzení:

$$\begin{aligned}
&\frac{x^{n+1}(x^2-1)}{x^{2n+2}+1}\frac{\cos(n+1)t}{x^2+1-2x\cos t} - \frac{x^n(x^2-1)}{x^{2n}+1}\frac{\cos nt}{x^2+1-2x\cos t} = \\
&= \frac{x^{n+1}(x^2-1)(x^{2n}+1)\cos(n+1)t - x^n(x^2-1)(x^{2n+2}+1)\cos nt}{(x^{2n}+1)(x^{2n+2}+1)(x^2+1-2x\cos t)} = \\
&= \frac{x^n(x^2-1)}{(x^{2n}+1)(x^{2n+2}+1)}\frac{(x^{2n}+1)x\cos(n+1)t - (x^{2n+2}+1)\cos nt}{x^2+1-2x\cos t} \overset{\text{Lemma 2}}{=} \\
&\overset{\text{Lemma 2}}{=} -\frac{x^n(x^2-1)}{(x^{2n}+1)(x^{2n+2}+1)}\left(-x^n + \sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt\right)
\end{aligned} \tag{A.8}$$

Lemma 4: Pro všechna $n \in \mathbb{N}$ a $x \in \langle 0;1)\,, t \in \mathbb{R}$ platí:

$$\int_0^\pi \frac{\cos nt}{1+x^2-2x\cos t}\mathrm{d}t = \frac{\pi x^n}{1-x^2} \tag{A.9}$$

Důkaz: Nejdříve ukážeme, že výraz pod zlomkovou čarou není nulový.

$$x^2 + 1 - 2x\cos t \geq x^2 + 1 - 2x = (x-1)^2 > 0 \tag{A.10}$$

Nyní k samotnému důkazu. Pokud $x = 0$, lemma dokážeme snadno:

$$\int_0^\pi \frac{\cos nt}{1+x^2-2x\cos t}\mathrm{d}t = \int_0^\pi \cos nt\,\mathrm{d}t = 0 = \frac{\pi x^n}{1-x^2} \tag{A.11}$$

Dále předpokládejme, že $x > 0$. Důkaz (indukcí podle n). Základní krok: pro $n = 1$ platí:

$$\begin{aligned}
&\int_0^\pi \frac{\cos t}{1+x^2-2x\cos t}\mathrm{d}t = \int_0^\pi \frac{2x\cos t}{2x(1+x^2-2x\cos t)}\mathrm{d}t = \\
&= \int_0^\pi \frac{x^2+1-x^2-1+2x\cos t}{2x(1+x^2-2x\cos t)}\mathrm{d}t = \int_0^\pi \left(\frac{x^2+1}{2x(1+x^2-2x\cos t)} - \frac{1}{2x}\right)\mathrm{d}t = \\
&= \frac{x^2+1}{x(x^2-1)}\int_0^\pi \frac{x^2+1}{2(1+x^2-2x\cos t)}\mathrm{d}t - \left[\frac{1}{2x}\right]_{t=0}^{\pi} \overset{\text{Lemma 1}}{=} \\
&\overset{\text{Lemma 1}}{=} \left[\frac{x^2+1}{x(x^2-1)}\operatorname{arctg}\left(\frac{x+1}{x-1}\operatorname{tg}\frac{t}{2}\right) - \frac{t}{2x}\right]_{t=0}^{\pi} = \\
&= -\frac{\pi(x^2+1)}{2x(x^2-1)} - \frac{\pi}{2x} = \frac{-\pi x^2 - \pi - \pi x^2 + \pi}{2x(x^2-1)} = \frac{\pi x}{1-x^2}
\end{aligned} \tag{A.12}$$

Poznámka: $\frac{x+1}{x-1}$ je záporné, proto $\lim_{t\to\pi^-}\operatorname{arctg}\left(\frac{x+1}{x-1}\operatorname{tg}\frac{t}{2}\right) = -\frac{\pi}{2}$.
Indukční krok: Předpokládejme, že lemma pro nějaké $n \in \mathbb{N}$ platí. Potom

z lemmatu 3 plyne:

$$\frac{x^{n+1}(x^2-1)}{x^{2n+2}+1}\frac{\cos(n+1)t}{x^2+1-2x\cos t} =$$
$$= \frac{x^n(x^2-1)}{x^{2n}+1}\frac{\cos nt}{x^2+1-2x\cos t} - \frac{x^n(x^2-1)}{(x^{2n}+1)(x^{2n+2}+1)}\left(-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt\right) \tag{A.13}$$

$$\frac{\cos(n+1)t}{x^2+1-2x\cos t} = \frac{x^{2n+2}+1}{x^{n+1}}\left(\frac{x^n}{x^{2n}+1}\frac{\cos nt}{x^2+1-2x\cos t} - \right.$$
$$\left. - \frac{x^n}{(x^{2n}+1)(x^{2n+2}+1)}(-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt)\right) =$$
$$= \frac{(x^{2n+2}+1)\frac{\cos nt}{x^2+1-2x\cos t} - (-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt)}{x(x^{2n}+1)} \tag{A.14}$$

A proto:

$$\int_0^\pi \frac{\cos(n+1)t}{1+x^2-2x\cos t}\mathrm{d}t =$$
$$= \int_0^\pi \frac{(x^{2n+2}+1)\frac{\cos nt}{x^2+1-2x\cos t} - (-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt)}{x(x^{2n}+1)}\mathrm{d}t =$$
$$= \frac{(x^{2n+2}+1)\int_0^\pi \frac{\cos nt}{x^2+1-2x\cos t}\mathrm{d}t - \int_0^\pi \left(-x^n+\sum_{i=0}^{n}(x^{2i}+1)x^{n-i}\cos nt\right)\mathrm{d}t}{x(x^{2n}+1)} =$$
$$= \frac{(x^{2n+2}+1)\int_0^\pi \frac{\cos nt}{x^2+1-2x\cos t}\mathrm{d}t - \int_0^\pi \left(-x^n+(x^0+1)x^{n-0}\cos 0t\right)\mathrm{d}t}{x(x^{2n}+1)} \overset{\text{I.P.}}{=}$$
$$\overset{\text{I.P.}}{=} \frac{(x^{2n+2}+1)\frac{\pi x^n}{1-x^2} - \pi x^n}{x(x^{2n}+1)} = \pi x^n \frac{x^{2n+2}+1-1+x^2}{x(x^{2n}+1)(1-x^2)} =$$
$$= \pi x^n \frac{x^2(x^{2n}+1)}{x(x^{2n}+1)(1-x^2)} = \frac{\pi x^{n+1}}{1-x^2} \tag{A.15}$$

A tedy lemma platí.

Tvrzení: Kruhové Cauchyovo rozdělení je pravděpodobnostní rozdělení.
Důkaz: Dokážeme, že integrál hustoty pravděpodobnosti je roven jedné. Otočíme si rozdělení tak, aby parametr μ byl kladné reálné číslo. Potom:

$$\int_{-\pi}^{\pi} f(\theta|\mu)\mathrm{d}\theta = \int_{-\pi}^{\pi} \frac{1-|\mu|^2}{2\pi}\cdot\frac{1}{|\,\mathrm{e}^{\theta i}-\mu|^2}\mathrm{d}\theta =$$
$$= \int_{-\pi}^{\pi} \frac{1-\mu^2}{2\pi((\cos\theta-\mu)^2-\sin^2\theta)}\mathrm{d}\theta = \int_{-\pi}^{\pi} \frac{1-\mu^2}{2\pi(1+\mu^2-2\mu\cos\theta)}\mathrm{d}\theta \overset{\text{Lemma 1}}{=}$$
$$\overset{\text{Lemma 1}}{=} -\frac{2}{\pi}\left[\operatorname{arctg}\left(\frac{\mu+1}{\mu-1}\operatorname{tg}\frac{\theta}{2}\right)\right]_{\theta=0}^{\pi} = \frac{2}{\pi}\cdot\frac{\pi}{2} = 1 \tag{A.16}$$

Poznámka: $\frac{\mu+1}{\mu-1}$ je záporné, a proto $\lim_{t\to\pi^-}\left(\frac{\mu+1}{\mu-1}\operatorname{tg}\frac{t}{2}\right) = -\frac{\pi}{2}$.
A tedy kruhové Cauchyovo rozdělení je pravděpodobnostní rozdělení.

Tvrzení: Distribuční funkce kruhového Cauchyova rozdělení pro $\mu \in \langle 0; 1)$ a $\theta \in (-\pi; \pi)$ je:

$$F_\mu(\theta) = \frac{1}{2} - \frac{1}{\pi} \operatorname{arctg}\left(\frac{\mu+1}{\mu-1} \operatorname{tg}\frac{\theta}{2}\right) \tag{A.17}$$

a kvantilová funkce je:

$$q_\mu(\alpha) = 2 \operatorname{arctg}\left(\frac{\mu-1}{\mu+1} \operatorname{tg}\left(\pi\left(\frac{1}{2} - \alpha\right)\right)\right) \tag{A.18}$$

Důkaz: Zintegrujeme hustotu pravděpodobnosti jako v důkazu předchozího tvrzení.

$$\begin{aligned}
&\int f(\theta|\mu)\mathrm{d}\theta = \int \frac{1-|\mu|^2}{2\pi} \cdot \frac{1}{|\,\mathrm{e}^{\theta i} - \mu|^2}\mathrm{d}\theta = \\
&= \int \frac{1-\mu^2}{2\pi((\cos\theta - \mu)^2 + \sin^2\theta)}\mathrm{d}\theta = \int \frac{1-\mu^2}{2\pi(1+\mu^2-2\mu\cos\theta)}\mathrm{d}\theta \overset{\text{Lemma 1}}{=} \\
&\overset{\text{Lemma 1}}{=} -\frac{1}{\pi} \operatorname{arctg}\left(\frac{\mu+1}{\mu-1} \operatorname{tg}\frac{\theta}{2}\right) + C
\end{aligned} \tag{A.19}$$

Následně konstantu C dopočítáme tak, aby $\lim_{\theta \to -\pi^+} F_\mu(\theta) = 0$. Snadno dojdeme k závěru, že $C = \frac{1}{2}$. Kvantilová funkce je inverzní funkcí k distribuční funkci, což vyjde:

$$q_\mu(\alpha) = 2 \operatorname{arctg}\left(\frac{\mu-1}{\mu+1} \operatorname{tg}\left(\pi\left(\frac{1}{2} - \alpha\right)\right)\right) \tag{A.20}$$

Tvrzení: Pro všechny momenty kruhového Cauchyova rozdělení platí:

$$m_n = \mu^n \tag{A.21}$$

Důkaz: Rozložíme parametr μ na absolutní hodnotu x a argument φ. Potom platí, že n-tý moment rozdělení otočeného o úhel φ je roven n-tému momentu otočenému o úhel $n\varphi$. Stačí se tedy zabývat reálnou částí rozdělení, pro které $\varphi = 0$. Převedeme absolutní hodnotu z komplexního čísla na funkci reálných čísel a spočítáme:

$$\begin{aligned}
&\int_{-\pi}^{\pi} \frac{1-x^2}{2\pi} \cdot \frac{\cos n\theta}{(\cos\theta - x)^2 + \sin^2\theta}\mathrm{d}\theta = \frac{1-x^2}{\pi}\int_0^{\pi} \frac{\cos n\theta}{(\cos\theta - x)^2 + \sin^2\theta}\mathrm{d}\theta \overset{\text{Lemma 4}}{=} \\
&\overset{\text{Lemma 4}}{=} \frac{1-x^2}{\pi} \cdot \frac{\pi x^n}{1-x^2} = x^n
\end{aligned} \tag{A.22}$$

A hodnota $|\mu|^n$ otočená o $n \arg\mu$ není nic jiného než μ^n.

Pokud chceme kruhové Cauchyovo rozdělení vnímat jako dvourozměrné rozdělení, můžeme vypočítat rozptyly a kovarianci následujícím způsobem. Nejdříve přepočítáme plošné momenty na kruhové momenty a potom dosadíme

podle výše uvedených vzorců. Nechť p_μ je pravděpodobnostní míra kruhového Cauchyova rozdělení s parametrem μ. Nejdříve předpokládejme, že μ je v intervalu $\langle 0; 1)$. Potom:

$$\begin{aligned}
DX = E(X^2) - (EX)^2 &= \int_\Gamma \cos^2\theta \,\mathrm{d}p_\mu - \left(\int_\Gamma \cos\theta \,\mathrm{d}p_\mu\right)^2 = \\
&= \int_\Gamma \left(\frac{1}{2} + \frac{1}{2}\cos 2\theta\right) \mathrm{d}p_\mu - \left(\int_\Gamma \cos\theta \,\mathrm{d}p_\mu\right)^2 = \\
&= \frac{1}{2} + \frac{1}{2}\int_\Gamma \cos 2\theta \,\mathrm{d}p_\mu - \left(\int_\Gamma \cos\theta \,\mathrm{d}p_\mu\right)^2 = \\
&= \frac{1}{2} + \frac{1}{2}\Re m_{2,\mu} - (\Re m_{1,\mu})^2 = \frac{1}{2} + \frac{1}{2}\mu^2 - \mu^2 = \frac{1-\mu^2}{2} \qquad \text{(A.23)} \\
DY = E(Y^2) - (EY)^2 &= \int_\Gamma \sin^2\theta \,\mathrm{d}p_\mu - \left(\int_\Gamma \sin\theta \,\mathrm{d}p_\mu\right)^2 = \\
&= \int_\Gamma \left(\frac{1}{2} - \frac{1}{2}\cos 2\theta\right) \mathrm{d}p_\mu - \left(\int_\Gamma \sin\theta \,\mathrm{d}p_\mu\right)^2 = \\
&= \frac{1}{2} - \frac{1}{2}\int_\Gamma \cos 2\theta \,\mathrm{d}p_\mu - \left(\int_\Gamma \sin\theta \,\mathrm{d}p_\mu\right)^2 = \\
&= \frac{1}{2} - \frac{1}{2}\Re m_{2,\mu} - (\Im m_{1,\mu})^2 = \frac{1}{2} - \frac{1}{2}\mu^2 - 0 = \frac{1-\mu^2}{2} \qquad \text{(A.24)} \\
\operatorname{cov}(X,Y) = E(XY) - (EX)(EY) &= \int_\Gamma \sin\theta\cos\theta \,\mathrm{d}p_\mu - \int_\Gamma \sin\theta \,\mathrm{d}p_\mu \int_\Gamma \cos\theta \,\mathrm{d}p_\mu = \\
&= \int_\Gamma \left(\frac{1}{2}\sin 2\theta\right) \mathrm{d}p_\mu - \int_\Gamma \sin\theta \,\mathrm{d}p_\mu \int_\Gamma \cos\theta \,\mathrm{d}p_\mu = \\
&= \frac{1}{2}\int_\Gamma \sin 2\theta \,\mathrm{d}p_\mu - \int_\Gamma \sin\theta \,\mathrm{d}p_\mu \int_\Gamma \cos\theta \,\mathrm{d}p_\mu = \\
&= \frac{1}{2}\Im m_{2,\mu} - \Im m_{1,\mu}\Re m_{1,\mu} = 0 \qquad \text{(A.25)}
\end{aligned}$$

Pokud je μ nějaké jiné číslo, výsledek pouze otočíme o $\arg\mu$. A protože kovarianční matice vyšla jako násobek jednotkové matice, můžeme psát obecně pro libovolné $|\mu| < 1$:

$$DX = \frac{1-|\mu|^2}{2} \qquad \text{(A.26)}$$

$$DY = \frac{1-|\mu|^2}{2} \qquad \text{(A.27)}$$

$$\operatorname{cov}(X,Y) = 0 \qquad \text{(A.28)}$$