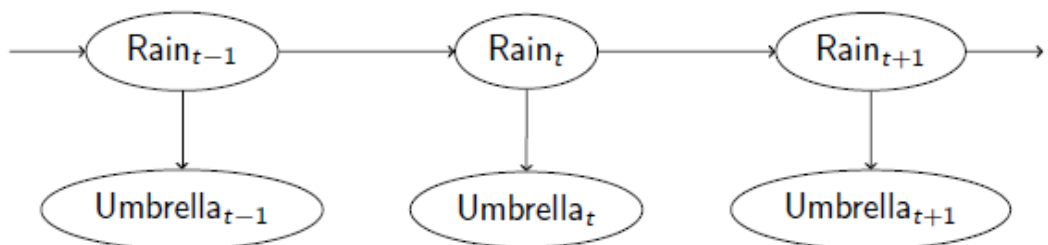


Gedächtnisprotokoll Pattern Analysis

- Prüfer: Dr.-Ing. Christian Rieß
- Zweitprüfer: hat sich nicht vorgestellt
- Uhrzeit& Ort: kamen am Tag vorher per Mail
- Stimmung: entspannt
- Benotung: mit PR vergleichbar 😊
- Zeit: verging wie im Fluge

1. Übersicht: PA-Wolke malen (siehe 2017-SS_PA_Summary.pdf)
2. Random Forests:
 - a. Wie funktioniert unsupervised learning mittels Density Estimation Tree?
 - b. Wie schaut ein Regression tree aus?
3. Density Estimation:
 - a. Parzen-Window-Estimator: Formel für $p(x)$ herleiten
 - b. Welche Alternative gibt es für die Density Estimation (s. Frage 2a)
4. Clustering:
 - a. Unterschiede zwischen k-means und mean-shift-algo (MSA) nennen, beide Algorithmen skizzieren
 - b. MSA:
 - i. Formel herleiten
 - ii. Welches Kernel verwendet man hier gerne? (Epanechnikov)
 - iii. Was passiert, wenn man nicht die 2-Norm, sondern eine andere Norm verwendet?
 - c. K-means:
 - i. Wie findet man ein gutes k? Mittels gap statistics (within-cluster-distance). Hier war es ihm besonders wichtig, dass die $W(C)$ -Kurve streng monoton fallend ist, auch bei dem „little dent“. Da wäre es in der Vorlesung von Vorteil gewesen, die Tafelzeichnung korrekt abzumalen.
 - ii. Was passiert bei Outlayers (Punkte, die weit weg vom Rest vom Fest liegen)?
5. HMM:
 - a. HMM malen (Grafik aus KI2 geklaut)



- b. Beide Assumptions (Markov, Output) hinschreiben
 - c. $p(x_1, \dots, x_N, y_1, \dots, y_N)$ vereinfachen: Indizes durch $i \in [1; N]$ ersetzen, Bayes' Theorem anwenden
 - d. Welche Algorithmen gibt es? (nur die Namen haben gereicht)
6. Manifold Learning: dazu kam leider keine Frage, obwohl MDS und ISOMAP so schön zu erklären gewesen wären 😞

1. PA Wolke aufmalen, kurze Erklärung zu den einzelnen Themen

2. Density estimation problem:

- Parzen window estimator (erklären, Formeln), parametric – nonparametric approach
- Wie wäre das Problem mit Random Forests realisiert (Erklärung, wie es funktioniert)
- Transferfrage: Was ist der Unterschied zwischen den beiden Ansätzen (Vor-, Nachteile)

Meine Antwort: Parzen Window estimator einfach zu berechnen, während der random forest trainiert werden muss (ggf. viele Bäume). Vorteil bei Wäldern, dass sie die Datendichte besser repräsentieren (Flexibilität), während der Parzen kernel nur eine feste Breite h hat.

3. Nach der sanften thematischen Überleitung sind wir nun bei den Random Forests angekommen

- Wie wird der Baum trainiert (Information gain maximiert, Entropie Formel $H(S)$ für den Spezialfall von density forests aufgeschrieben).

4. HMMs:

- Welche Annahme ist grundlegend für die graphischen Modelle: Markov Annahme, für HMM und MRF erklärt
- Hier wurde ich nach "Inferenzproblem" im HMM gefragt, was mich etwas verwirrt hat. Gemeint war das "Wortproblem" $P(o_1, \dots, o_T | \lambda)$ und das Problem der Rekonstruktion der Zustandskette $\max_S P(O, S | \lambda)$
- Trainingsproblem: Baum-Welch-Formeln, das war nicht mehr schwer
- Baum-Welch Formeln skizziert, Viterbi-Algorithmus vollständig aufgeschrieben
- Zu welcher Klasse von Algorithmen gehört Baum-Welch: Gewollt war Expectation-Maximization-Algorithmus (EM), meine beiden Antworten waren greedy und approximativ, was nicht falsch ist aber nicht unmittelbar zum Fach gehörend
- Wenn er greedy ist, dann landet man häufig auf lokalen Maxima, wie kann man dem Problem abhelfen? Leichte Transferfrage: Mehrmals mit unterschiedlichen Initialisierungen laufen lassen und die Performance jeweils auf den Testdaten prüfen
- Wie waren die b_j (zum Zustand S_j gehörige Wahrscheinlichkeitsverteilung) in der Übung: kontinuierlich (Wahrscheinlichkeitsdichte) statt diskret

Die Atmosphäre war sehr entspannt. Vorbereitet hatte ich mich mit den Vorlesungsvideos aus dem Vorjahr, da ich die Vorlesung aus zeitlichen Gründen nicht besuchen konnte. Außerdem habe ich die angegebene Literatur benutzt, dies war sehr hilfreich zum Verständnis der Formeln.

FSI INFORMATIK

- [Wiki](#)
- [Forum](#)
- [Chat](#)
- [Videos](#)
- [Evaluationen](#)

- [Startseite](#)
- [Studienstart](#)
- [Prüfungen](#)
- [Studium](#)
- [Deine FSI](#)
- [Hochschulpolitik](#)
- [Kontakt](#)

Sie befinden sich hier: [Termine](#) » [Prüfungsfragen und Altklausuren](#) » [Hauptstudiumsprüfungen](#) » [Lehrstuhl 5](#) » [pa-2019-07-29](#) ([Übersicht](#))

Pattern Analysis Prüfer: Riess

Note: 1,3

Insgesamt wären noch ein paar Tage besser um letzte Lücken zu stopfen und evtl. sogar ein paar Herleitungen nachvollziehen zu können.

Alle meine Materialien: <https://github.com/reinerzer/pattern-analysis-ss19>

Q: Give an overview over the lecture A: Find structure in data:

1. Density Estimation
2. Clustering
3. Segmentation (Trees)
4. Hidden Data

Dann jeweils bisschen in die Tiefe gegangen, unterbrochen bei K-Means

Q: K-Means is considered „hard “, GMM clustering considered „soft “ - why? A: K-Means yields a strict assignment of a data point to a cluster, GMM yields assignments to all clusters with probabilities

Q: ... how do we select a K? A: Three ways: Tibshirani Gap Staticstics [Beschreibung] ... unterbrochen

Q: How do we select the number of GMMs? A: I suppose it would also work with a weighted within-cluster distance gap like in tibshirani but we did it with CRP / Gibbs sampling [Beschreibung]

Q: In that sense what is the „rich get richer “ factor here? A: The prior for a Gaussian [Beschreibung]

Q: For manifold learning what kinds of algorithms do we have A: Common element: similarity matrix and projection on low-dim space via eigenvectors of the similarity matrix. Explain MDS and Laplacian [unterbrochen, bevor ich zu den anderen kommen konnte]

Q: How does Laplacian Eigenmaps work? A:

1. Construct Adjacency Graph
2. Compute Affinities
3. Eigendecomposition
4. Lower-Dimensional Embedding

Q: What is the objective function of the Laplacian Eigenmaps? A: $\sum \|x_i - x_j\|^2 w_{ij}$ (oder so)

Q: Yes but then if we derive it what do we have? A: [falsch hingeschrieben, gesucht war $D^{-1}L * x = \lambda * x$] and then we can do Eigendecomposition

Q: When do we have an advantage of Density Trees and when is a heat kernel preferable Die Frage habe ich nicht genau verstanden, hab aber dann ein bisschen erklärt wie die weights in einem Density Tree segmentierten Bereich mit den Kovarianzen der leaf densities der Form des Gaussians rechnung tragen. Das hat ihm gefallen ...

Q: Without going into detail of HMMs and MRF - what are graphical models and what do they model? A: Statistical Dependencies between random variables.

Q: Please formularize that A: [nicht ganz richtig gemacht, was er als größtes Problem an der Prüfung gesehen hat] Independency assumptions formal hinschreiben (Produkt über alle variablen usw...) Er meinte, dass es sehr wichtig ist das sicher zu können.

Q: What is a MRF? What do we use it for? A: We try to model structures in „grid-like “ data by trying to infer hidden information based on the observable grid. Observations only depend on the hidden counterpart. hidden counterparts only depend on the observable counterpart and the grid neighbors ... [mehr beschreibung, überführung in GRF, Submodularity condition]

Q: Please state the GRF probability A: $p(x) = 1/Z * e^{-H(x)}$; $H(x) = \sum V_m(x)$

Q: Please state an example potential function A: They depend on what you actually try to infer in the data (denoising, segmentation). For denoising we have for the pairwise potentials $\|f_{ij} - f'\|^2$

Vorbereitung: zu zweit 6 Tage, Zusammenfassung aus verschiedenen Skripten (da wir nur die VL 2018 gesehen haben) geschrieben. Meta-Summary mit den wichtigsten Formeln (4 Seiten) zum auswendig Lernen (hilft beim Verständnis und bei einer live demo wenn man nicht zögert beim schreiben) erstellt.

- [Zeige Quelltext](#)
- [Ältere Versionen](#) [Links hierher](#)
- ...
- [Letzte Änderungen](#)
- [Anmelden](#)
- [Datenschutz](#)

FSI INFORMATIK

- [Wiki](#)
- [Forum](#)
- [Chat](#)
- [Videos](#)
- [Evaluationen](#)
- [Startseite](#)
- [Studienstart](#)
- [Prüfungen](#)
- [Studium](#)
- [Deine FSI](#)
- [Hochschulpolitik](#)
- [Kontakt](#)

Sie befinden sich hier: [Termine](#) » [Prüfungsfragen und Altklausuren](#) » [Hauptstudiumsprüfungen](#) » [Lehrstuhl 5](#) » [pa-2019-08-20](#) ([Übersicht](#))

Grade: 1.0

The exam started off with Prof. Riess asking to give an overview of all topics in the course that we discussed.

Q. He asked me about density estimation and what the idea behind that was.

A. I started with what the goal behind density estimation is, the approaches we could follow for the task (parametric, non-parametric). He stopped me to ask what parametric and non-parametric meant. I gave an explanation and detailed the approaches we could follow for each of the above (ML/MAP estimate for the former, Parzen window estimate for the latter).

Q. Yeah the bare bones Parzen window estimate. Could you detail on that perhaps.

A. I started once again with the idea behind using a kernel function to estimate the data, and briefly discussed how it's formulated by writing a bunch of formulas.

Q. Great. So we implemented something like this in the first exercise where we were trying to visualize the results of using this approach. Say you have a friend that sends you an output image of a raccoon but you don't know whether they used a Gaussian or a box kernel. Is there a way to know this? He then drew a 1D feature space such that it had samples on the x-axis.

A. So I assume that with a Gaussian kernel, you'd have a smoother output visually but you might get block artifacts from a box kernel. He was clearly looking for something more, and gave me a bit of time with it, asking me what 'smoothing' would mean from a visual perspective, but then I drew a box-like pdf as a result of sliding a box kernel through the samples. For the Gaussian kernel, I drew a bell-shaped estimated pdf, and this was what he was looking for.

Q. So we also discussed clustering. What is the idea behind that?

A. I explained what the goal with a clustering algorithm is, and then briefly touched upon the two algorithms that we generally use (K-means and a GMM).

Q. We also discussed another algorithm in class that isn't a conventional clustering algorithm but can be used to perform the task.

A. Yes, we talked about the mean shift algorithm. I explained the idea behind the mean shift (how we are looking for modes of density), and how we can get to these modes by looking for the zero gradients of the density. I then explained how the mean shift works for a clustering algorithm.

Q. Say your friend sends you results from another project he is working on, and its basically an image of a dataset where some clustering has been performed. But your friend isn't too talkative and you want to find out how he might have arrived at the clusters.

A. I drew two cases where a conventional k-means algorithm doesn't work well (oblong data/unequal variance and unequal samples in clusters). K-means tends to produce unintuitive results in such cases, but mean shift would work. If the data has spherical clusters it perhaps wouldn't be possible to judge which algorithm the friend used by seeing the visual results. But for the former datasets, a judgement is possible.

Q. I would like to move on to a probabilistic graphical model. But before we go into details, tell me how we tackle the probability space in these problems? He wrote down a joint probability density function (similar to the lecture), with N input observations and M output observations.

A. I remarked how using the entire combinations leads to an $N \times M$ probabilistic space which perhaps is impossible to model. So we use the Naive Bayes assumption. I then wrote a bunch of formulas detailing how the original space could be broken down with this assumption and the model becomes tractable. Also drew the probabilistic graph for Naive Bayes.

Q. Great, but the Naive Bayes is just a little naive. Could you list some potential application-specific problems, and how we address them?

A. For data where sequence of events is important (listed an example from a reference: words and parts-of-speech as labels, the sequence information is important). We can use a Hidden Markov Model, wrote a few formulas to show how the Naive Bayes could be adapted for HMM.

Q. Could you draw a HMM graph?

A. Listed the elements of HMM, drew the graph, explained what a left-right HMM is.

Q. Great, we also have some tasks in HMM and specific algorithms for those tasks.

A. Explained Inference and Training, and the tasks and algorithms in each.

Q. Could you explain what happens in the training phase? You don't need to write the formulas.

A. Gave the algorithm name, its an iterative expectation maximization approach. Explained what we do in expectation, and what we do in maximization.

Q. So the solution isn't globally optimal?

A. I said it wasn't, to which he asked how we could get to a globally optimal solution? Perhaps use different initialization schemes and then perform cross-validation to choose most optimal solution.

Environment: The atmosphere was quite friendly. The entire exam was more of a discussion rather than a monologue. After I gave him a general overview of the topic, he had specific questions where we went into greater detail.

Preparation: What helps is to have an intuitive understanding of each topic, so you can explain what a particular concept means, after which you can go into the details. I used the notes from the class. I went back to Google in case I had trouble reproducing the intuition behind a certain technique. Five or six days before the exam is a good enough time to prepare if you have gone through 60 or 70% of the material once before. I created small web diagrams for each topic, and wrote the formulas for all of them. Revised using these one day before the exam. The references are a good source for developing a well-rounded understanding of the topics. Went through around 70% of the references that were compulsory, mostly skipped the additional ones.

- [Zeige Quelltext](#)
- [Ältere Versionen Links hierher](#)
- ...
- [Letzte Änderungen](#)
- [Anmelden](#)
- [Datenschutz](#)

Pattern Analysis, SS 2017

Dr. Riess is very friendly and the atmosphere of the examination is pretty relaxed, so no need to worry.

1 Intro

At the beginning I had to present the usual big picture. It's important to illustrate the relationship between all the different topics (for example: Parzen-window \rightarrow Mean shift \rightarrow We can use this for clustering \rightarrow Clustering (k-means, k-NN, GMM) \rightarrow we need to specify number of clusters \rightarrow Dirichlet process as an alternative and so on...)

2 Clustering

How does k-means work?

1. Random assignment or best guess
2. Assign samples to cluster with nearest cluster mean
3. Recompute means
4. Iterate 2) 3) until convergence

Specifying k isn't so easy in most cases, is there any heuristic we can use?

Compute $J = \sum_{c \in C} \sum_{x_i \in c} \|x_i - \mu_c\|_2^2$ for different k's and search for "knee" in resulting graph.

Can we improve this procedure?

Also compute J on uniformly distributed samples. Search for k where $J_{\text{uniform}} - J_{\text{training set}}$ is biggest. (You can look this heuristic up in *Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Section 12*)

Ok, now we can go a different way and do soft clustering via GMMs without having to specify the exact number of clusters.

I told him everything he said about Dirichlet Processes (Distribution over distributions, CRP as a way to constructively "sample" from a Dirichlet distribution, why Chinese Restaurant, how does the algorithm look like, how to compute the affinities, ...).

Now we have those affinities. How do we compute which cluster we should assign a sample to?

Normalize sum of affinities to one \rightarrow We now have a discrete pdf and are able to randomly draw a cluster assignment from it.

How do we draw samples from a one dimensional pdf?

Compute cdf, draw uniformly distributed number v between 0 and 1. Our sample x is the one where $\text{cdf}(x) = v$.

Imagine you'd like to segment a picture via clustering. How would you do that?

Transform each pixel into feature vector (x, y, r, g, b) , apply clustering algorithm. Important: scale dimensions.

When you use Mean Shift for clustering, do you have to specify the number of clusters?

Not directly, but the kernel size influences the number of maxima we get. I then had to illustrate one a 1d sample distributing which local maxima mean shift would detect with different window sizes.

Transition to new topic: *Imagine your working for a company and your boss says, well that's nice and all... But all these clustering algorithms are very 90s. I want you to do image segmentation via MRFs. How could we do that?*

3 MRF

I first explained how a typical MRF would look like (grid structure), mentioned the Hammersly-Clifford theorem and wrote out the general pdf factorization of a Gibbs random field ($p(x) = \frac{1}{Z} e^{-\sum_c H_c(x_c)}$). After that I told him how the binary and pairwise potentials would look like if we would do image denoising.

Now, how would the potentials/energy functions look like if we would do segmentation?

This was all a bit fuzzy. We would assume smoothness of the values of the random variables (0 means pixel is part of background, 1 means pixel is part of foreground), so the pairwise potentials would be something like the difference between two neighboring random variables. For the unary potentials he said that this is a bit harder and its ok that I don't know how to specify these.

Imagine for the moment we would have fully specified our energy functions, how can we minimize our overall energy function?

I began with the submodularity condition, every energy function fulfilling this condition is graph representable, this means we can minimize the function by finding the minimum s-t-cut.

How would the subgraph for a unary potential look like?

Showed him the same thing he showed us in the lecture.

Now thats cool and all, but why bother constructing this graph and minimizing the function via minimum cut?

There exist algorithms which run in polynomial runtime.

I made some lecture notes, which you can find here (there may be some errors and some lectures are missing):

<https://github.com/elbundy/SS17/tree/master/PA>

Prüfungsprotokoll Pattern Analysis SS18

Angenehme Atmosphäre. Entspannteste meiner mündlichen Prüfungen bisher.

- PA Wolke aufzeichnen
- Was ist das allgemeine Density Estimation Problem?
 - o Aus einem finiten Set von Beobachtungen eine kontinuierliche Wahrscheinlichkeitsverteilung ableiten die den Beobachtungen zugrunde liegt.
- In PR wurden parametric density estimation Lösungen gezeigt, hier non-parametric. Einer davon war der Parzen Window estimator, wie funktioniert der?
 - o Anstatt einen Histogram-based Ansatz zur Abschätzung der Wahrscheinlichkeitsverteilung zu wählen o.Ä. wird ein Kernel Window über die Datenpunkte geschoben.
 - o An jeder Position eines Datenpunktes wird der Kernel ausgewertet.
 - o Die schlussendliche Wahrscheinlichkeitsverteilung erhält man über Summierung aller Window-Auswertungen.
- Parzen Window gibt es schon sehr lange. Die Idee wurde dann später nochmals aufgegriffen. Wo?
 - o Mean Shift Algorithmus
- Was ist die Idee des Mean Shift Algorithmus?
 - o Man benutzt den Ansatz der Parzen Window Estimation um eine Wahrscheinlichkeitsverteilung abzuschätzen. (Name Mean Shift kommt davon, dass wenn man den Epanechnikov Kernel benutzt, der Kernel den Mean der Punkte in seinem Einzugsbereich berechnet)
 - o Danach sucht man die Maxima der Verteilung indem man Nullstellen des Gradienten sucht.
 - o Von jedem Datenpunkt aus wird iterativ ein gradient ascent mit Hilfe der Mean Shift Gleichung durchgeführt bis man am Maximum angekommen ist
- Wenn der Mean Shift zum Clustern benutzt wird, wie wird dabei entschieden welcher Datenpunkt zu welchem Cluster gehört?
 - o Jedes Maximum bildet einen eigenen Cluster
 - o Jeder Punkt der beim gradient ascent zu einem Maximum führt gehört zu dessen Cluster
- Wenn der Mean Shift zum Clustern auf Bilddaten angewandt wird, wie funktioniert das?
 - o Da habe ich gemeint dass die Farben des Bildes geclustert werden und deshalb die Pixelfarbe als dritte Dimension neben x- & y-Koodrinat kodiert wird
 - ➔ Er hat gemeint dass die Farben kodieren sehr ambitioniert ist, aber mit Graustufen klappt das
- Wenn man jetzt Mean Shift mit normalem k-Means vergleicht, was ist da so der Unterschied?
 - o Ich war bei der Frage etwas blank dagestanden
 - o Habe dann gemeint dass bei z.B. nicht-konvexen Daten Probleme beim k-Means entstehen könnten -> Habe drei konzentrische Kreise gemalt, Punkte auf einem Kreis sollen ein Cluster bilden, und beschrieben dass k-Means dabei probleme hätte, habe aber auch gesagt dass ich mir nicht sicher bin warum der Mean Shift dabei besser funktionieren sollte
 - o Er fragte dann wie ich das Problem so „hindefinieren“ könnte damit der Mean Shift gut funktioniert

- Ich meinte dass wenn man lokale Maxima entlang der Ringe definiert würde der Mean Shift diese finden und korrekt clustern.
- Er war damit anscheinend zufrieden und ergänzte noch dass man auch eine einzige lange Gradientensteigung entlang der Ringe definieren könnte denen der Mean Shift dann folgt bis zum Maximum des entsprechenden Ringes.
- Bei Clustering haben wir auch noch über spectral clustering geredet. Wie funktioniert das?
 - Man betrachtet die Affinitäten der Datenpunkte
 - ➔ -> Was sind Affinitäten?
 - Inverse einer Distanz
 - Man bildet den Graph Laplacian mit den Affinitäten und erhält durch dessen Eigenvektoren einen neuen Koordinatenraum wo k-Means angewandt wird.
- Der Graph Laplacian kommt auch bei den Laplacian Eigenmaps vor. Was ist die Idee davon?
 - „Konzept von Laplacian Eigenmaps ist dass nahe Punkte in hoher Dimension auch in niedriger Dimension nahe sind.“
 - Habe dann das ganze Konzept stückweise beschrieben und auch einige Formeln zur Herleitung des Graph Laplacian aufgeschrieben.
 - Noch kurz beschrieben wie man mit den Eigenvektoren dann ein embedding in eine niedrigere Dimension machen kann.
- Themawechsel: Hidden Markov Models, was ist das? Können Sie eins aufmalen?
 - Graphisches Modell zur Modellierung von 1D sequentiellen Daten (zeitliche Daten)
 - Habe ein Modell mit 4 States aufgemalt und alle nötigen Pfade eingezeichnet
- Was hat das Markov im Namen zu bedeuten?
 - Das kommt von der „markovian assumption“ die besagt dass die Wahrscheinlichkeit dafür in einem State zu sein nur vom State eins davor abhängt und nicht von anderen.
 - Habe dann noch gemeint dass das wünschenswert ist um die Abhängigkeiten in den Gleichungen für die Wahrscheinlichkeiten viel kleiner zu bekommen.
- Man braucht 3 Algorithmen um mit HMMs zu arbeiten. Welche?
 - Forward/Backward, Viterbi und Baum-Welch
 - Habe dann noch die Formeln für die alpha Variable im Forward-Algorithmus aufgeschrieben und ein paar zum Training mit den Baum-Welch Gleichungen.
 - Er hat zwischendurch noch genau nach der ξ -Variable bei Baum-Welch gefragt und ich habe beschrieben dass die Wahrscheinlichkeit für exakt einen State-Übergang von s_i nach s_j beschreibt mit den Wahrscheinlichkeiten bis dahin und von da weg enthalten.

Note: 1,0

FSI INFORMATIK

- [Wiki](#)
- [Forum](#)
- [Chat](#)
- [Videos](#)
- [Evaluations](#)
- [Home page](#)
- [Start of studies](#)
- [exams](#)
- [studies](#)
- [Your FSI](#)
- [University policy](#)
- [Contact](#)

You are here: [Events](#) » [Exam questions and old exams](#) » [Main course exams](#) » [Chair 5](#) » [Exam questions / procedure](#) ([Overview](#))

- Table of contents

- [Exam questions / procedure](#)
- [proceed](#)
- [the atmosphere](#)
- [Learning process](#)

Examiner: Dr. Riess, assessor:? Result: very good

Exam questions / procedure

Q: Let's start then, what have we done in the VL?

A: (PA cloud painted on)

Density Estimation

Q: Okay, huh ... Density Estimation, what do we need that for?

A: (Just started with "lecture") Discrete observations, but want P.D.F for statistical key figures or to generate new samples. $p_R = (\text{samples within} / \text{samples total}) * VR$, formula for $p(x) = 1 / N$ sum kernel (x_i, x) painted, "Hypercube" kernel explained, Gauss explained, influence of Sigma / kernel size on P.D.F , shown what it looks like (step-like with Hypercube → not differentiable), underfitting / overfitting discussed, determination of the parameter via ML estimation

Q: How do we now know whether our parameter makes sense (exact question no longer in mind)

A: Cross-validation, we take samples out of our original set and then see whether they can be explained well by our found P.D.F.

Density Forests

Q: Now we had also talked about density forests. How do they work?

A: Would you start with the basic principle of Random Forests, if that's ok, for your understanding?

Q: Ok

A: Briefly discussed general trees: Decision tree structure narrated, IG maximize - but now back to the Density Forests: Unsupervised problem, i.e. no classes, we fit a Gauss into each leaf, information gain is drawn, formula briefly explained: $|\text{cov}(\Sigma)|$ is a kind of volume function, $|S_j|$ the number of samples in the respective branch, $|S_j|$ We need all of them as a measure in order not to optimize on a single sample, we then want to minimize $H(S_j)$ as much as possible because we want to subtract as little as possible in order to maximize the IG. Use blurring to smoothen, otherwise the gaussians 'break off' very brutally and are not continuous.

Q: Ok, how should our $|\text{cov}(\Sigma)|$ then really look like if we want to maximize IG?

A: (used briefly), painted the logarithm \rightarrow I had a hang for a moment, painted the wrong curve (basics ...), talked about it a bit

Q: Now again to be sure: How should our Sigma be - as compact or as large as possible?

A: As compact as possible?

Q: bingo.

(Note: the student before me said he was asked what the difference between density trees and estimation is - with trees, each leaf has its own covariance matrix, which means that different sigmas can be used, which is accordingly better)

HMM

Q: But now I want to do something else. HMM, what is it, what do you need it for?

A: Spam example brought: 0-1 vector of all words, the individual dimensions are statistically dependent \rightarrow is stupid because you can't handle it \rightarrow for example "Viagra" and "buy" more likely than something else \rightarrow We introduce Markov Assumption \rightarrow then use HMMs as a kind of automaton \rightarrow Markov Assumption says that every state has a P.D.F. that is not dependent on previous states \rightarrow now you can deal with it

Q: Okay, now you say "statistically dependent", the mathematician would say too ...

A: (At first I didn't quite know what he was trying to get out of, but then I thought it could only be the formula, I wrote it down): $p(x | \text{all others}) \leftarrow$ somehow roughly painted, hadn't learned it by heart.

Q: fits ...

A: Jo, dann haben wir unsere HMM, die schaut so aus (irgendeine HMM-Struktur hingemalt), da haben wir $\lambda = \pi, A, B$, kurz erklärt was das jeweils ist. Jetzt gehen wir mal nochmal zu nem anderen Beispiel, Wort erkennen: Wir wollen also wissen ob das ein bestimmtes Wort ist. Dann haben wir drei „Probleme“: Wahrscheinlichkeit dass es ein bestimmtes Wort ist, der wahrscheinlichste Pfad, Training.

(Kurze Pause gelassen, kam keine Rückfrage, also einfach weiter gemacht...)

Eigentlich ist das ja ganz logisch wie wahrscheinlich eine gewisse Sequenz ist → „nichtoptimierte“ Variante mit den ganzen Summenzeichen am Anfang hin gemalt → das ist aber halt doof, weil sehr aufwändig.

Q: Was heißt sehr aufwändig?

A: Naja, $O(m^n)$ oder $O(n^m)$... muss ich kurz nachdenken, Sekunde, hab das nicht auswendig gelernt da logisch... nachgedacht → das richtige gesagt

Q: Genau, und was heißt „sehr aufwändig“ fachlich gesprochen?

A: Exponentieller Aufwand.

Q: Jup

A: Ja, also auf jeden Fall kann man das optimieren → dynamic programming → Forward und Backward Algorithm angesprochen → Forward erklärt.

Q: Ok, da möchte ich jetzt dann noch die Formeln sehen

A: Grob hingeschrieben

Q: Jap.

A: Backward weiter erklärt, dann zum Viterbi übergegangen, Viterbi erklärt → statt Summe nur noch max, Pfad merken.. Naja, ist ja alles schön und ugt, jetzt halt die Frage wie trainiert man das Ganze... Baum-Welch angesprochen, Formeln hingeschrieben, erklärt

Q: Passt, dann ist die Zeit rum, dann gehen sie doch mal raus...

Vorgehen

Kurz ins Thema einführen, dann kurz Pause lassen um zu schauen ob Rückfragen kommen, sonst einfach weiter reden. Wenn Fragen kommen darauf schauen, dass man die möglichst präzise beantwortet. Danach konnte ich eigentlich immer einfach den „Vortragsmodus“ weiter durchziehen.

Atmosphäre

Schön entspannt, super nett, reitet nicht auf Kleinigkeiten rum. Denke wenn man ihm zeigt dass man das Zeug verstanden hat passt das. Bei der O-Notation und Baum-Welch hatte ich das Zeug nicht auswendig gelernt sondern gesagt „Sekunde, brauche kurz um das logisch zu rekonstruieren“ → dann erklärt und er gibt einem da auch die Zeit ohne hektisch zu werden.

Lernvorgehen

Durch VL-Notizen gegangen, VL-Notizen zusammengefasst, Grundidee gelernt, ggf. „Hauptformeln“ bzw. Optimierungsprobleme, Verständnisfragen ergoogelt.

Herleitungen habe ich ausgeklammert und immer die Kernaussage mitgenommen (nach dem Motto „mit 0 gleichsetzen, auflösen, dann kommt diese Formel raus: X). Prüfung ging aber glücklicherweise eher auf das Algorithmisch-Statistische statt Mathematische ein, daher weiß ich nicht, ob das funktioniert (gerade bei den Punkten Graph Laplacian/Eigendecomposition etc.).

Er hatte eine Reihe an „Must Read “-papers und „Optional “-Papers zur Verfügung gestellt. Die „Must Read “ hatte ich die Kernpunkte gelesen und den Rest überflogen. Wenn man da an der ein oder anderen Stelle Bezug drauf nimmt (Density Forests: Kurz Vergleich zu GMMs gebracht, dass die Forests halt noch „Uncertainty “-Informationen haben) kommt das denke ich ganz gut. Optional-Papers maximal überflogen.

- [Zeige Quelltext](#)
- [Older versions link here](#)
- ...
- [Letzte Änderungen](#)
- [Anmelden](#)
- [Datenschutz](#)

FSI INFORMATIK

- [Wiki](#)
- [Forum](#)
- [Chat](#)
- [Videos](#)
- [Evaluations](#)
- [Home page](#)
- [Start of studies](#)
- [exams](#)
- [studies](#)
- [Your FSI](#)
- [University policy](#)
- [Contact](#)

You are here: [Events](#) » [Exam questions and old exams](#) » [Main course exams](#) » [Chair 5](#) » [Generally](#) ([Overview](#))

- Table of contents

- [Generally](#)
- [ask](#)
- [preparation](#)
- [General examination & assessment](#)

Generally

The exam was taken in the office of Prof Freiling, with an advisor. In general, a very relaxed atmosphere that you can step into. Mr. Riess likes to start with a topic overview and then continue with a random topic. Riess is more likely to let you talk and digs when he wants to know something.

ask

Riess : What kind of topics did we deal with?

Roughly going through the course of the lecture and naming the methods we discussed for each topic

Riess : What are graphical models?

Graphical Models explains → Markov assumption, state only depends on direct predecessor

Riess : Write down the formula for the Markov assumption.

Here I had a bit of a blockage, despite my support (via Bayes etc.) I couldn't get this formula right down ...

Riess : Markov Random Fields: How do they work?

The structure of MRFs explained, hidden variables + observations, transformation in Gibbs Random Field by Hammersley – Clifford theorem → formula written down and explained → briefly on unary and pairwise potentials, but here without formulas

Riess : Let's move on to another area. How does the mean-shift algorithm work?

Goal: Find extremes of the $P.D.F.$ → the derivation of the Parzen-Window formula roughly explained → Update formula written down. (Here he wanted to know why the Epanechnikov kernel is so good)

Riess : The k-means algorithm would be similar. Briefly how does it work?

Briefly explains the steps we take there ...

Riess : (Here he drew me a sketch with two cluster centers) Can it happen with mean-shift or k-means that I cluster a point to the left of the left cluster center in my right center?

Here I thought out loud for a while and discussed with him until I had the solution. With k-means it cannot happen because in principle there would only be a border between the two centers and everything on the left comes to the left center and everything on the right to the right ... But with mean shift we do a gradient-ascent → if now in a curve of our point to the left of the left center around the left center more and more points land in the kernel window, our kernel window moves to the right center

Riess : Finally, what steps are we taking at Spectral Clustering?

The 6 steps from the lecture roughly explained.

preparation

Man musste alle Übungen bestanden haben, um überhaupt die Möglichkeit zu haben, die Prüfung anerkannt zu bekommen. War auch an sich gut, die Übungen gemacht zu haben, vielleicht auch weil man sich viel dafür selbst erarbeiten musste. Ansonsten alle Vorlesungsvideos angeschaut und die Mitschriften noch einmal zusammengefasst. Die Videos waren auch nötig da es leider kein Skript oder Vorlesungsfolien gibt. Es gab auch ein paar Paper die man lesen sollte, die habe ich zu ca 70% gelesen, teilweise hat es sehr geholfen die Methoden besser zu verstehen, da es meist etwas ausführlicher war. Manche Paper waren aber auch nahezu gar nicht hilfreich.

Prüfung Allgemein & Bewertung

A very relaxed and calm atmosphere during the exam. You don't feel pressured at all. Mr Riess tries to help if you are upset, but you definitely have time to think before he helps. In itself a fair test with a fair evaluation. What bothered him the most was my problem with the Markov assumption that I needed some discussion during the sketch in order to come up with the solution, which was not bad.

- [Show source code](#)
- [Older versions link here](#)
- ...
- [last changes](#)
- [Register](#)
- [data protection](#)

Pa Gedächtnisprotokoll

MRF:

Prinzip erklären

Zeichnung des Modells (aus der Vorlesung)

Erklärung der Potentiale (unary / pairwise) und Verknüpfungen

i.e. unary potential: kann Gauss Funktion sein, Funktion aus Hidden Variable und Observation oder pairwise könnte eine Funktion die geringsten Abstand zwischen den Labels misst.

Graph cuts am Beispiel der unary potential zeigen (wie in der Vorlesung schon dargestellt)

Formel: $P(1,1) + P(0,0) > P(1,0) + P(0,1)$ ("gleich gelabelt grösser als "ungleich" gelabelte Potentiale) darstellen

Dirichlet Process (zur Bestimmung der Anzahl der Cluster ?...)

Erklären ("Verteilung von Verteilungen", "sampling einer Verteilung", $DP(\alpha, G)$, etc.)

Stickbreaking Process: Formel (für Gewichte) + Grafisch darstellen

Formel darstellen wie "Wahrscheinlichkeitsvektor" aus den Gewichten und einer Indikatorfunktion gebildet werden

Anschließen erklären wie man aus den "Wahrscheinlichkeitsvektor" gesamplet wird

Chinese Restaurant Process: grafisch darstellen und erklären

Gibbs sampling: eben der Algorithmus des ganzen - Formel für Gauss darstellen, mit dem

Konzentrationsparameter α + eben auf die Option einen neuen Tisch zu kreieren (da braucht man ja dann die Priors für den Mean und der Varianz des Gauss)

Mean Shift Algorithmus

An Beispiel (2D graph mit Punkten) erklären

Man schiebt eben ein Fenster (quadratisch bzw für Epanechnikov kreisförmig) bis zum Maximum

Was passiert mit der Windowsize – an 1D Beispiel (aus der Vorlesung bekannt) gezeigt

Gedächtnisprotokoll Pattern Analysis

Datum: 10.10.17

Prüfer: Dr. Christian Riess

- Zuerst sollte ich die übliche PA Wolke aufmalen
 - Hab ich natürlich gemacht und alle Themengebiete erläutert
- Dann wollte er, dass ich die verschiedenen Methoden für Clustering erkläre
 - Da habe ich zuerst mit Mean Shift angefangen. Die Formel erklärt, den Epanechnikov Kernel erwähnt und dass der mit Abgeleitet dasselbe ist wie einfach nur den Mean berechnen, daher der Name.
 - Als nächstes habe ich K-means erklärt, aber nur anhand einer Grafik und ohne Formeln (also in unserem Semester haben wir auch keine Formeln in der Vorlesung dafür besprochen)
 - Als letzte habe ich noch Chinese Restaurant Process erklärt. Da eben die Formeln für die Affinitäten und den Stick-Breaking Process. Weiterhin habe ich eben erwähnt, dass man sich hier nicht explizit für eine feste Anzahl K an Clustern entscheidet, sondern das ganze indirekt über den expansion Parameter alpha steuert.
- Ihm war daraufhin noch wichtig, dass ich auch erwähne wie es nach den Affinitäten weitergeht.
 - Dann habe ich erklärt, dass man die Affinitäten in eine Liste schreibt und diese Liste auf 1 normalisiert und dann zufällig von dieser Liste sampelt.
- Er wollte dann wissen wie man von einer PDF sampelt
 - Habe ihm aufgezeichnet, dass man die CDF berechnet und dann uniform die Y-Achse sampelt.
- Danach sollte ich noch erklären wie man bei k-means auf das richtige K kommt
 - Indem man die Withincluster distance vs. die Clusteranzahl K plottet und nach einem Signifikanten Knick sucht. Weiterhin könnte man auch noch die Funktion für das Clustern auf einer uniformen Verteilung mit einzeichnen und dann den Unterschied der beiden Funktionen anschauen.
- Themenwechsel. Ich sollte ihm die Idee hinter Random Forests erklären
 - Man hat einen Baum (weak learner) und trainiert den zufällig auf dem Datensatz. Dabei habe ich halt noch die verschiedenen split functions und den Information Gain erwähnt. Am Ende werden alle Bäume geaveraged und das eben als Ergebnis genommen.
- Er wollte dann Möglichkeiten für die Randomisierung von den Bäumen wissen
 - Als erstes kann man nur ein subset des Trainingsdatensatzes nehmen
 - Oder die Splitting function parameter dem Zufall überlassen (also quasi bei den hyperplanes (in 2D) die Steigung und den Offset)
 - Hab dann noch gesagt, dass man theoretisch auch die Art der Splitting function dem Zufall überlassen kann, das aber nicht so üblich ist.
- Dann sind wir zum Thema HMM übergegangen. Hier sollte ich einfach mal erklären was ein HMM ist.
 - Naja ein HMM ist ein graphical model um sequentielle Daten zu modellieren. Habe ihm dann ein einfaches Model aufgezeichnet und die verschiedenen Variablen eingezeichnet (A , B , π). Habe noch erwähnt, dass man es meistens für Spracherkennung benutzt. Habe dann noch die 3 grundlegenden

Probleme: Evaluation, Decoding und Training mit den jeweiligen Ansätzen dafür (Forward/Backward algorithm, Viterbi Algorithm, Baum-Welch Algorithm) genannt

- Ob ich den Forward/Backward Algorithmus genauer erklären könnte
 - Habe ihm dann die 3 Schritte hingeschrieben und erklärt
- Dann sollte ich noch kurz den Trainingsprozess erklären, ohne Formeln
 - Habe dann eben die verschiedenen Schritte des Baum-Welch Algorithmus erwähnt und gesagt wofür die Variablen stehen, bzw. was man damit berechnet. Z.B. war ihm wichtig, dass man für die Berechnung des Gammas Alpha und Beta aus dem FW/BW Algorithmus braucht und dass das Gamma ja quasi alle Pfade hin und weg von einer observation sind.

Zusammenfassung: Ja ganz entspannte Prüfungsatmosphäre und faire Fragen. Ich habe sehr viele Formeln hin geschrieben, ich weiß nicht in wie weit das Nötig gewesen wäre. Ich fand die Fragen waren alle völlig im Rahmen (also nichts was nicht in der Vorlesung dran kam). Ihm war aber schon wichtig, dass man es verstanden hatte und man nicht nur die Formeln abgeschrieben hat. Meine Note war auf **sehr** zufriedenstellend.

Zur Vorbereitung habe ich alle Vorlesungsvideos angeschaut und davon eine Zusammenfassung geschrieben. Es lohnt sich auch die Referenzen die er auf der Webseite (www5.cs.fau.de/.....) angibt durchzulesen. Das hilft oft das Thema besser zu verstehen. Insbesondere für das Thema Random Forests fand ich das eine gute Hilfe, da wir da teilweise nicht so viel (Formeln) aufgeschrieben hatten.

Prüfungsprotokoll Pattern Analysis SS 17

Prüfer: Riess

Beisitzer: kannte ich nicht

Atmosphäre: Prüfer hatte 20 Minuten Verspätung, dann war es wie ein Gespräch über die Themen - relativ angenehm, bin aber immer sehr nervös.

1. Einstieg mit der PA-Wolke
2. Hidden Markov Model
 - (a) was ist es und warum ist es toll (Spracherkennung, state sequence muss nicht bekannt sein)
 - (b) drei Algorithmen nennen (Forward/Backward , Baum-Welsh, Viterbi)
 - (c) Model erklären (mit Formeln)
 - (d) Training mit Formeln herleiten
3. Clustering
 - (a) welche Möglichkeiten gibt es (Mean-Shift, CRP, K-Means/Soft K-Means)
 - (b) welche Methode ist für welches Problem gut nutzbar / Unterschiede
 - (c) CRP erklärt mit Formeln

bei diesem Thema war Transfer gefragt. Hab manchmal nicht die Antworten aus der Vorlesung geben können, weil ich mich nicht erinnern konnte. Habe aber dann einen anderen Ansatz überlegt, was ihm sehr gut gefallen hat.

Note: 1.3

1 Pattern Analysis
2 4.10.2016
3 Prüfer: Christian Riess
4
5 Sehr nette Atmosphäre.
6 Prüfung mit Übung (7,5 ECTS)
7
8 Folgende Themen wurden in der Prüfung behandelt
9
10 Kurzer Überblick über alle Themen geben (Themen nennen)
11 Clustering:
12 zwei Möglichkeiten nennen, bei denen keine Clusteranzahlvorgabe gemacht wird
13 -> MDS, Chinese Restaurant Problem (CRP)
14 MDS: Konkrete Herleitung mit Formeln von MDS, wie genau bekomme ich
15 die Lösung (was ist mein X), wie kann ich die Dimension von meinem
16 Raum reduzieren?
17 CRP: Formeln für die Wahrscheinlichkeit der einzelnen Cluster? Wie
18 ergibt sich diese insgesamt k+1 Wahrscheinlichkeiten? Bei k Clustern
19 bekommen wir k Wahrscheinlichkeiten + eine Wahrscheinlichkeit für ein
20 neues Cluster, was ist genau μ_{00} bei der Formel für eine neue WSK?
21 Wie weiß ich einen Datenpunkt ein Cluster zu? $r \leq \text{cdf}(k)$ - wobei r
22 eine gleichverteilte Zufallsvariable zw 0 und 1 ist und k das Cluster.
23 Male hierzu eine Beispiels pdf und cdf und veranschauliche.
24 Was ist das Alpha bei einem DP und wie wirkt es sich auf die
25 Wahrscheinlichkeiten aus?
26
27 Manifold Learning:
28 Welche Methoden gibt es?
29 LLE: konkrete Herleitung mit Formeln und einzelnen Schritten, wie
30 komme ich zu meiner Lösung
31
32 Aus Zeitgründen keine Fragen zu HHM, wobei das bei der Studentin auch
33 konkret gefragt wurde.
34
35 und noch fragen zu "seinem Lieblingsthema" MRF:
36 Zeichne und erkläre MRF.
37 Graph cuts -> Psi muss eine submodular Function sein. Formel für
38 Bedingung! Was bedeutet das konkret für Psi? -> Psi muss als Funktion
39 so gewählt werden, dass es der Ungleichung genügt!
40
41
42 Keine Fragen zu der Übung, das einzige was ich an "Code" schreiben
43 musste, war wie ich ein Cluster bei dem CRP einem Datenpunkt zuweise:
44 `r = math.rand()`
45 `for k in range(len(cdf)):`
46 `if r <= cdf(k):`
47 `# weise x zu cluster k zu`
48
49 Mit Benotung sehr zufrieden!