

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](#)