

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.