

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