

Abgabe: 18.11.16 - 10 Uhr

Bitte Bearbeiten Sie den Zettel in Zweier- oder Dreiergruppen.

Das Abgabeformat sollte ein pdf (mit einer Beschreibung, was gemacht wurde, mit eventuell relevanten Codeauszügen und Plots) und eine Menge von Code Textdateien (Für jede Aufgabe eine Datei) sein.

Aufgabe 1: Expectation Maximization - 6 Punkte

Schreiben Sie ein Programm, dass den Datensatz 2d-em.csv einliest und den Expectation Maximization Algorithmus darauf ausführt. Variieren Sie dabei die Anzahl der Cluster und bewerten Sie die Güte der Annäherung gegenüber der Klassenzahl (Nutzen Sie dafür eine Metrik aus der Vorlesung bzw. dem Tutorium oder beschreiben Sie, für welche Clustermengen Sie gute Ergebnisse erzielt haben bzw. für welche weniger gute). Plotten Sie dabei für einige (ausgewählte) Iteration die Samples und die Cluster als Ellipsen. Hinweis: Sie können die Ellipse eines Clusters als Kreis zeichnen, dessen Punkte mit der Kovarianzmatrix eines Clusters transformiert wurden.

Aufgabe 2: Expectation Maximization II - 4 Punkte

Führen Sie Expectation Maximization auf allen Daten der [digits](http://statweb.stanford.edu/tibs/ElemStatLearn/datasets/zip.train.gz)¹ aus und geben Sie die Konfusionsmatrix aus. Beschreiben Sie außerdem, wie man die Einträge in der Konfusionsmatrix interpretieren kann.

¹<http://statweb.stanford.edu/tibs/ElemStatLearn/datasets/zip.train.gz>