

Blatt 2

Trainings und Testmenge sind wiederum die pendigits Daten aus dem UCI Machine Learning Repository:

<http://archive.ics.uci.edu/ml/machine-learning-databases/pendigits/>

Aufgabe 1: Visualisierung

Stelle die Mittelwerte der Klassen für die Trainingsmenge graphisch dar (als Linienzuege der Ziffern).

(10 Punkte)

Aufgabe 2: Klassifikation

Modelliere die Klassen mit der multivariaten Normalverteilung und weise jedem Sample aus der Testmenge diejenige Klasse zu, die am wahrscheinlichsten ist. Du kannst, musst aber keine apriori Wahrscheinlichkeiten (priors) verwenden, da diese in unserem Datensatz ungefähr gleich sind.

Evaluere den Klassifikator: Wie viele Samples aus der Testmenge klassifiziert der Klassifikator korrekt? Wie sieht die Confusion Matrix aus?

http://en.wikipedia.org/wiki/Confusion_matrix

(15 Punkte)