

# 5. Übungszettel Mustererkennung WS15/16

Prof. Raúl Rojas, Fritz Ulbrich  
Institut für Informatik, Freie Universität Berlin  
Abgabe Online bis Mittwoch, 02.12.15, 10 Uhr

Bitte laden Sie ihre Lösung der Aufgaben als **pdf-Datei** hoch.  
Quellcode können Sie optional als Archiv anhängen.

## 1. Aufgabe (5 Punkte): Lineare Regression

Importieren Sie den Datensatz **fish.txt** in matlab. Das Format ist:

index ; the age of the fish ; the water temperature in degrees Celsius ; the length of the fish

- a. (3 Punkte) Schätzen Sie den Wert für "length" anhand der Parameter "age" und "temperature" mit linearer Regression. Visualisieren Sie dreidimensional die tatsächlichen Datenpunkte, die geschätzten Datenpunkte, sowie die Abstände der tatsächlichen zu den geschätzten Datenpunkten.
- b. (2 Punkte) Visualisieren Sie die durch die in Aufgabe a) berechneten Koeffizienten definierte Ebene.

## 2. Aufgabe (5 Punkte): Subset Selection

Importieren Sie den Datensatz **winequality-red.txt** in matlab. Das Format ist:

fixed acidity; volatile acidity; citric acid; residual sugar; chlorides; free sulfur dioxide; total sulfur dioxide; density; pH; sulphates; alcohol; quality (score between 0 and 10)

Schätzen Sie den Wert für "quality" mit linearer Regression anhand aller möglichen Kombinationen der anderen Parameter (also jeweils für alle Einer-, Zweier-, Dreierkombinationen, usw.) und berechnen jeweils die Summe der quadratischen Abweichungen zwischen den geschätzten und tatsächlichen Werten für "quality". Visualisieren Sie dies als zweidimensionalen Plot. Auf der x-Achse steht dabei die Anzahl der verwendeten Parameter, auf der y-Achse die Summe der quadratischen Abweichungen (für alle Kombinationen der jeweiligen Anzahl der Parameter) .