

8. Übungszettel Mustererkennung WS15/16

Prof. Raúl Rojas, Fritz Ulbrich
Institut für Informatik, Freie Universität Berlin
Abgabe Online bis Mittwoch, 13.01.16, 10 Uhr

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

1. Aufgabe (10 Punkte): Logistische Regression

Importieren Sie den Datensatz **spiders.txt**. Der Datensatz beschreibt das Vorkommen der "burrowing wolf spider" (*Lycosa ishikariana*) an verschiedenen Stränden in Abhängigkeit von der Größe der Sandkörner (Format: grain size (mm), spiders (0=absent, 1=present)).

- (5 Punkte) Implementieren Sie das in der Vorlesung vorgestellte Verfahren (Gradientenabstieg) zur Bestimmung der Regressionskoeffizienten β für die logistische Regression. Verwenden Sie dabei als initialen Wert $\beta=[0;0]$. Verwenden Sie die Schrittweite $\alpha=0.1$ für den Gradientenabstieg. Plotten Sie jeweils nach 1, 10, 100 und 1000 Iterationen die Funktion $p(x, \beta)$, sowie die Datenpunkte.
- (1 Punkt) Berechnen Sie, ab welcher Größe der Sandkörner es wahrscheinlicher ist, Spinnen anzutreffen (als keine Spinnen anzutreffen).
- (4 Punkte) Plotten Sie die log-likelihood Funktion $l(\beta)$ des Datensatzes für das Intervall $\beta_0 = [-100, 100]$ und $\beta_1 = [-100, 100]$ in einem dreidimensionalen Koordinatensystem. Plotten Sie in dem selben Diagramm die log-likelihood des von Ihnen in Aufgabe a) berechneten β als dreidimensionalen Punkt.
(Siehe http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/tutorials/logistic.pdf für die Definition der log-likelihood Funktion).