Übung zur VL Mustererkennung
Wintersemester 2012/2013

# Blatt 5

Die Daten zu diesem Aufgabenblatt sind die gleichen wie in den vorhergehenden Übungen:

1. Prostate Data aus Hastie, Tibshirani, Friedman
   http://www-stat.stanford.edu/~tibs/ElemStatLearn/
2. Pendigits Data vom UCI Machine Learning
   http://archive.ics.uci.edu/ml/machine-learning-databases/pendigits/

## Aufgabe 1: Ridge Regression

Verwende die Daten unter 1. (Prostate Data). Normiere die Merkmale: ziehe den Mittelwert ab und teile durch die Standardabweichung, so dass die Merkmale in Einheiten von Standardabweichungen vorliegen. Implementiere Ridge Regression und bestimme die Koeffizienten $\alpha_1, \alpha_2..\alpha_8$. Trage die Werte von $\alpha_i, i = 1..8$ gegen $df(\lambda)$ auf. $df(\lambda)$ sind die effektiven Freiheitsgrade, sie sollten im Intervall [0,8] liegen. Die Formel für $df$ findest Du im Buch von Hastie et al. auf Seite 68 (corrected 5th printing, Feb. 2011). Der gesuchte Plot ist zum Verlgleich auch dort (Figure 3.8). Notiz: $\alpha$ aus der Vorlesung heisst im Buch von Hastie et al. $\beta$.

(10 Punkte)

## Aufgabe 2: Bootstrap

Nimm die besten Merkmale aus dem Subset der Größe 3 aus der letzten Übung. Wähle aus den Daten (Prostate Data) 50 Samples (Werte von Menschen mit Prostatakrebs) per Zufall aus, mit Zurücklegen. Bestimme die Werte der Koeffizienten mit Linearer Regression. Führe das Experiment insgesamt 100 mal durch, merke die Koeffizienten in jedem Schritt und berechne $\mu \pm 2\sigma$ für jeden Koeffizienten.

(10 Punkte)

## Aufgabe 3: Experiment

Bestimme die Kovarianzmatrix $\Sigma$ für eine Ziffer Deiner Wahl aus den Pendigits Trainingsdaten (wie in Übung 2). Mache nun folgendes Experiment: Wähle zufällig einen Vektor $\mathbf{x_0}$ der Länge 16. Multipliziere ihn mit $\Sigma$ um den Vektor $\mathbf{x_1}$ zu erhalten. Nach k Schritten erhält man $\mathbf{x_k} = \Sigma^k \mathbf{x_0}$. Was beobachtest Du für die Richtung von $\mathbf{x_k}$? (Eventuell musst Du die Vektoren zwischen den Schritten auf die Länge 1 normieren, damit die Werte nicht zu groß werden.)

(5 Punkte)