Übung zur VL Mustererkennung
Wintersemester 2012/2013

Blatt 4

Die Daten zu diesem Aufgabenblatt findet Ihr unter http://www-stat.stanford.edu/~tibs/ElemStatLearn/ unter Prostate Data. In beiden Aufgaben geht es um lineare Regression.

Aufgabe 1: Prostate Cancer

Im Datensatz gibt es 8 Parameter

`lcavol, lweight, age, lbph, svi, lcp, gleason und pgg45.`

Verwende alle Parameter, um den Wert für `lpsa` vorherzusagen und bestimme die Summe der quadratischen Abweichungen. Verwende als Trainingsdaten die Daten, die mit T markiert sind. Die mit F markierten Daten sind die Testdaten.

(5 Punkte)

Aufgabe 2: Subset Selection

Versuche nun die Parameter zu reduzieren. Berechne dazu die Summe der quadratischen Abweichungen für alle Kombinationen der Parameter (für alle 28 Zweierkombinationen, alle 56 Dreierkombinationen, usw.). Stelle die gefundenen Subsets als Graph dar. Auf der x-Achse sollte die Anzahl der Parameter im Subset stehen und auf der y-Achse sollte die Summe der quadratischen Abweichungen stehen.

(15 Punkte)