

## Aufgabenblatt 6

Abgabe bis zum 14.12.2007

Neben der schriftlichen Abgabe, sind die Programmieraufgaben **zusätzlich** per E-Mail an den Tutor zu schicken. Eine E-Mail ersetzt nicht die schriftliche Abgabe.

### Aufgabe 1 (21 Punkte)

- Laden Sie sich die in der Vorlesung beschriebenen Daten für die Vorhersage des prostate specific antigen (psa) von <http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/prostate.data> herunter (das Datenformat wird im Tutorium besprochen).
- (7 Pkt.) Implementieren Sie die Methode der linearen Regression mittels der Pseudoinversen für die Vorhersage des lpsa (log of psa) aus den obigen experimentellen Daten.
- (7 Pkt.) Subset Selection. Berechnen Sie den quadratischen Fehler für die lineare Approximation von lpsa für alle Untermengen der features und plotten Sie den Fehler über der Größe der Untermenge.
- (7 Pkt.) Bootstrapping. Teilen Sie zufällig die Daten in Trainings- ( $\frac{2}{3}$ ) und Testmenge ( $\frac{1}{3}$ ) und berechnen Sie die Subset Selection. Wieder holen Sie dies 30 mal. Berechnen Sie Mittelwert und Standardabweichung des Fehlers und stellen Sie dies als Fehlerkurve mit Konfidenzintervall (für ein  $\sigma$ ) dar.