

In dieser Übung geht es darum die PCA- bzw. PLS-Regression durchzuführen und mit den Ergebnissen der Shrinkage-Methoden zu vergleichen. Wir verwenden den Datensatz *prostate* aus dem Paket *ElemStatLearn*.

A 1

- a) Erstelle eine Korrelationsmatrix für die Variablen *UV lcavol*, *lweight*, *age*, *lbph*, *svi*, *lcp*, *gleason* und *pgg45*. Welche Variablen korrelieren stark miteinander? Interpretiere!
- b) Führe eine PCA-Regression mit den *UV lcavol*, *lweight*, *age*, *lbph*, *svi*, *lcp*, *gleason* und *pgg45* und der AV *lpsa* durch (*pcr* aus dem Paket *pls*). Wähle die Optionen so, dass die Variablen vor der PCA-Regression standardisiert werden, sowie der Fehler einer 10-fachen Cross-Validation für jede Anzahl von Hauptkomponenten verwendet wird.
- c) Vergleiche die Fehler der Cross-Validations und gib die Anzahl mit dem kleinsten Fehler an.
- d) Erstelle einen Plot der MSE-Werte für die jeweilige Hauptkomponentenanzahl und wähle eine optimale Anzahl (Begründe!).
- e) Führe mit der in d) gefunden Anzahl eine einfache Cross-Validation durch (verwende die Variable *train*) und vergleiche den MSE-Wert mit den beiden MSE-Werten der Shrinkage Methoden bei gleicher Cross-Validation (letztes Übungsblatt).

A 2

- a) Führe eine PLS-Regression mit den *UV lcavol*, *lweight*, *age*, *lbph*, *svi*, *lcp*, *gleason* und *pgg45* und der AV *lpsa* durch (*plsr* aus dem Paket *pls*). Wähle die Optionen so, dass die Variablen vor der PLS-Regression standardisiert werden, sowie der Fehler einer 10-fachen Cross-Validation für jede Anzahl von Hauptkomponenten verwendet wird.
- b) Vergleiche die Fehler der Cross-Validations und gib die Anzahl mit dem kleinsten Fehler an.
- c) Erstelle einen Plot der MSE-Werte für die jeweilige Hauptkomponentenanzahl und wähle eine optimale Anzahl (Begründe!).
- d) Führe mit der in c) gefunden Anzahl eine einfache Cross-Validation durch (verwende die Variable *train*) und vergleiche den MSE-Wert mit den beiden MSE-Werten der Shrinkage Methoden sowie mit dem Wert aus Aufgabe 1.