

In diesem Übungsblatt wird der Datensatz *prostate* aus dem Paket *ElemStatLearn* verwendet. Hier wird der Einfluss verschiedener Variablen von Prostata-Krebs-Patienten auf deren PSA-Wert untersucht (nähere Variablenbeschreibung siehe Vorlesung).

A 1 In dieser Aufgabe soll die Ridge-Regression auf den Datensatz angewendet werden.

a) Führe eine lineare Regression mit den UV *lcavol*, *lweight*, *age*, *lbph*, *svi*, *lcp*, *gleason* und *pgg45* und der AV *lpsa* durch und speichere die Koeffizienten in einem Vektor.

b) Führe eine Ridge-Regression durch und vergleiche die Koeffizientenschätzer für $\lambda = 0$ und $\lambda = 10$ mit den Schätzern aus a). Interpretiere die Ergebnisse.

Hinweis: Die Ridge-Regression wird mit der Funktion *glmnet* (Paket *glmnet*) mit der Option $\alpha=0$ durchgeführt. Diese Funktion benötigt als Input die Modell Matrix der linearen Regression (Funktion *model.matrix*), sowie den Vektor der Werte der AV (*lpsa*).

c) Erzeuge einen Plot der Koeffizienten für verschiedene λ -Werte.

d) Führe 'per Hand' jeweils eine einfache Cross-Validation für $\lambda = 0$, $\lambda = 0.09$ und $\lambda = 2$ durch und vergleiche die MSE Schätzer.

e) Ermittle mithilfe der Funktion *cv.glmnet* den optimalen Schätzer für λ . Erzeuge einen Plot für den Test-MSE-Schätzer abhängig von λ . Vergleiche das Ergebnis mit d).

f) Vergleiche die Koeffizienten aus der Ridge-Regression mit optimalem λ mit den Koeffizienten aus a).

A 2 In dieser Aufgabe soll das Lasso-Verfahren auf den Datensatz angewendet werden.

a) Führe Lasso-Verfahren durch und vergleiche die Koeffizientenschätzer für $\lambda = 0$ und $\lambda = 10$ mit den gefundenen Schätzern aus A1 a). Interpretiere die Ergebnisse.

Hinweis: Das Lasso-Verfahren wird mit der Funktion *glmnet* mit der Option $\alpha=1$ durchgeführt.

b) Erzeuge einen Plot der Koeffizienten für verschiedene λ -Werte.

c) Führe 'per Hand' jeweils eine einfache Cross-Validation für $\lambda = 0$, $\lambda = 0.002$ und $\lambda = 1$ durch und vergleiche die MSE Schätzer.

d) Ermittle mithilfe der Funktion *cv.glmnet* den optimalen Schätzer für λ . Erzeuge einen Plot für den Test-MSE-Schätzer abhängig von λ . Vergleiche das Ergebnis mit c).

e) Vergleiche die Koeffizienten aus der Ridge-Regression mit optimalem λ mit den Koeffizienten aus A1 a). Interpretiere die Ergebnisse.