

Aufgaben zur Vorlesung
Multivariate Verfahren
Übungsblatt X

1. Beantworten Sie folgende Fragen bzw. bearbeiten Sie folgende Arbeitsanweisungen:
 - (a) Erläutern Sie, was bei einer Kreuzvalidierung geschieht.
 - (b) Was ist der bias-variance-tradeoff?
 - (c) Erklären Sie, warum man mit Hilfe von LASSO eine Variablenselektion durchführen kann.
 - (d) Was versteht man unter einer Matrix-Regularisierung?
2. Sie führen eine lineare Regression auf einem p -dimensionalen Datensatz durch und erhalten als Koeffizientenschätzer den Vektor $\hat{\beta}_{OLS}$. Die Annahmen der linearen Regression sind wie im Skript.
 - (a) Zeigen Sie: $\hat{\beta}_{OLS}$ ist erwartungstreu.
 - (b) Berechnen Sie die Varianz von $\hat{\beta}_{OLS}$.
3. Bei der Arbeit für ein großes Industrieunternehmen haben Ihre Kollegen und Sie in den letzten drei Monaten Daten für eine Prozessoptimierung gesammelt. Die gesammelten Daten sind im Vektor Y (abhängige Variable) und in der Matrix X (Regressoren) festgehalten:

$$Y = \begin{pmatrix} 2\sqrt{2} \\ \sqrt{3} \\ \sqrt{6} \end{pmatrix}, X = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{pmatrix}.$$

Sie wollen zunächst eine lineare Regression durchführen. Jedoch sind momentan Ihre Computersysteme auf unbestimmte Zeit ausgefallen und Sie benötigen die Ergebnisse bis heute Abend.

- (a) Zeigen Sie, dass Ihre Regressordaten orthogonal sind.
- (b) Berechnen Sie den geschätzten Koeffizientenvektor für die lineare Regression.

Nach der Präsentation der Ergebnisse vor dem Abteilungsleiter, kommt dieser auf die Idee eine Ridge Regression durchzuführen, da er den Begriff schonmal gehört hat. Er möchte, dass Sie $t = 10$ wählen.

- (c) Wie lautet die Formel für den Koeffizientenvektor einer Ridge Regression im nicht orthogonalen Fall?
- (d) Berechnen Sie λ für den vorliegenden Fall. Um wieviel Prozent werden die OLS Koeffizienten bei der Ridge Regression geschrumpft?

- (e) Wie lautet der geschätzte Koeffizientenvektor $\hat{\beta}_R$ der Ridge Regression?
4. Ein Freund von Ihnen hat für eine Seminararbeit eine lineare Regression durchgeführt. Sein geschätzter Koeffizientenvektor ist $\hat{\beta}_{OLS} = (1, -2, 3, -4)'$. Er möchte nun LASSO anwenden, um die Anzahl der Variablen zu verringern. Dazu möchte er $t = 12$ benutzen.
- (a) Ist seine Wahl für t sinnvoll?
- (b) Zu welchem Vorgehen raten Sie ihm?
5. Rechnerübung: Laden Sie den Datensatz *imports* aus Ilias herunter. Dieser enthält Daten zu importierten Autos. Führen Sie eine LASSO Regression für den logarithmierten Preis (1. Spalte im Datensatz) auf die restlichen Variablen des Datensatzes durch. Welche Regressoren würden Sie verwenden?