

Universität zu Köln
Mathematisches Institut



Seminararbeit zum Thema:
Lineare Regression and Regularisierung

vorgelegt von
Aylin Demirel
Matrikelnummer: 7312414

Dozent: Zoran Nikolić

17. Dezember 2020

Abbildungsverzeichnis

1.1	Darstellung einer Geraden $Y = \beta_0 + \beta_1 X$	2
-----	---	---

Abkürzungsverzeichnis

LR	Lineare Regression
ELR	Einfache lineare Regression
MLR	Multiple lineare Regression

1 Lineare Regression

Die lineare Regression ist ein statistisches Verfahren, bzw. eine überwachte Lernmethode, zur Analyse des Einflusses von einer oder mehreren unabhängigen Variablen auf eine metrische abhängige Variable.

Sofern eine solche Analyse den Zusammenhang zwischen zwei Zufallsvariablen behandelt, spricht man von einer einfachen linearen Regression (ELR). Handelt es sich um die Analyse eines Zusammenhangs zwischen mehr als zwei Zufallsvariablen, ist generell von einer multiplen linearen Regression (MLR) die Rede.

1.1 Einfache lineare Regression

Bei der ELR wird eine abhängige Variable lediglich durch eine unabhängige Variable erklärt. Das Modell geht daher von zwei metrischen Größen aus: einer Einflussgröße X , die unabhängige Variable, und einer Zielgröße Y , die abhängige Variable.

Wir wollen Y als Funktion von X darstellen. Im einfachsten Fall liegen alle Punkte auf einer Geraden, somit wird ein solcher Zusammenhang durch die lineare Funktion

$$Y = \beta_0 + \beta_1 X \quad (1.1)$$

dargestellt, wobei β_0 und β_1 die Regressionskoeffizienten sind. Genauer genommen ist β_0 die Regressionskonstante (engl. *Intercept*) und β_1 das Regressionsgewicht (engl. *Slope*). In Abbildung 1.1 erkennt man solch ein lineares Modell, wobei alle Punkte auf der Geraden liegen und es keine Abweichungen zwischen den tatsächlichen Werten y_i und den vorhergesagten Werten \hat{y}_i gibt.

In der Praxis werden jedoch bei der Vorhersage von Y durch X Fehler gemacht. Wir definieren $\epsilon_i = y_i - \hat{y}_i$ als Fehlerterm, der durch die Abweichung zwischen beobachteten Werten y_i und vorhergesagten Werten \hat{y}_i geschätzt wird. Daraus ergibt sich folgendes Modell:

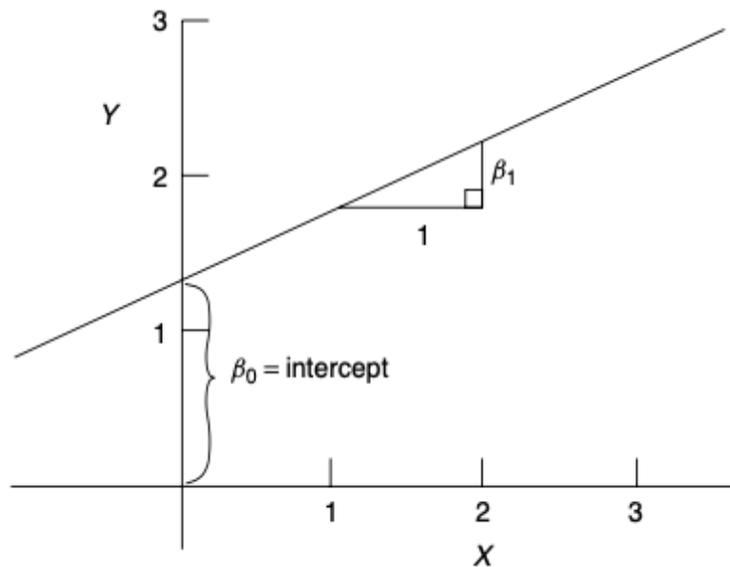
$$Y = \beta_0 + \beta_1 X + \epsilon \quad (1.2)$$

Wir nutzen das Symbol $\hat{\cdot}$, um den geschätzten Wert bzw. vorhergesagten für einen unbekanntem Parameter oder Koeffizienten zu bezeichnen.

1.1.1 Schätzung der Regressionskoeffizienten

β_0 und β_1 sind unbekannt und müssen anhand von Daten geschätzt werden. Die Daten $(x_1, y_1), \dots, (x_n, y_n)$ repräsentieren n Beobachtungspaare, von denen jedes aus einer Messung von X und einer Messung von Y besteht. Das Ziel ist es, Koeffizientenschätzungen β_0 und β_1 zu erhalten, so dass das lineare Modell (1.2) die Daten gut erklärt und eine passende Vorhersage bieten kann. Visuell gesehen bedeutet dies, dass die Gerade so nah wie möglich an den Datenpunkten liegen soll.

Es gibt eine Reihe von Möglichkeiten, die Nähe der Geraden zu den Datenpunkten zu messen. Der bei weitem häufigste Ansatz besteht in der Minimierung des Kriteriums der kleinsten Quadrate und diesen Ansatz verfolgen wir im weiteren Verlauf.

Abbildung 1.1: Darstellung einer Geraden $Y = \beta_0 + \beta_1 X$ [3, S. 20]

Methode der kleinsten Quadrate nach Gauß

Sei $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ für $i = 1, \dots, n$ die Vorhersage für Y basierend auf dem i -ten Wert von X . Dann stellt $\epsilon_i = y_i - \hat{y}_i$ das sogenannte *Residuum* bzw. den *Fehlerterm* dar. Dies ist die Differenz des tatsächlich beobachteten Wertes y_i und des durch unser Modell vorhergesagten Wertes \hat{y}_i . Das Ziel ist es, eine Regressionsgerade zu ermitteln, die nah an den gegebenen Datenpunkten liegt, um so auch eine möglichst „gute“ Vorhersage treffen zu können. Um dies zu erreichen, möchte man selbstverständlich die Fehler ϵ_i minimieren, damit die Abstände zwischen der Geraden und den Datenpunkten minimal sind. Es reicht jedoch nicht aus, lediglich die Summe aller Residuen zu minimieren, da sich positive und negative ϵ_i ausgleichen könnten, sodass die Summe annähernd oder gleich 0 ist. Daher wurde die *Residuenquadratsumme* definiert, die wie folgt beschrieben ist:

$$\begin{aligned}
 RSS &= \sum_{i=1}^n \epsilon_i^2 \\
 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 &= (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + \dots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2
 \end{aligned} \tag{1.3}$$

Die Methode der kleinsten Quadrate wählt die Regressionskoeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$, so dass das RSS minimiert wird. Wird die Gleichung aus (1.3) nach $\hat{\beta}_1$ abgeleitet, erhält

man die Minimierer:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.4)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad (1.5)$$

wobei $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ und $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ die Mittelwerte der Datenpunkte sind. Diese Minimierer definieren die Koeffizientenschätzungen der kleinsten Quadrate für die ELR.

Varianzen der Kleinste-Quadrate-Schätzer

Um zu beurteilen, wie gut unsere Kleinste-Quadrate-Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ die tatsächlichen Regressionskoeffizienten β_0 und β_1 widerspiegeln, berechnen wir deren Varianzen mit $\sigma^2 = \text{Var}[\epsilon]$:

$$SE(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \quad (1.6)$$

$$SE(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

σ ist ein unbekannter Parameter, der geschätzt werden kann mit $RSE = \sqrt{\frac{1}{n-2}RSS}$.

Konfidenzintervalle der Kleinste-Quadrate-Schätzer

Die Varianzen können zur Berechnung von Konfidenzintervallen verwendet werden. Ein 95%-Konfidenzintervall ist als ein Wertebereich definiert, der mit einer Wahrscheinlichkeit von 95% den wahren unbekanntem Wert des Parameters enthält. Der Bereich wird in Form von unteren und oberen Grenzen definiert, die aus der Datenstichprobe berechnet werden. Für ELR hat das 95%-Konfidenzintervall für β_0 und β_1 die Form

$$\begin{aligned} &\hat{\beta}_0 \pm 2 \cdot SE(\hat{\beta}_0) \\ &\hat{\beta}_1 \pm 2 \cdot SE(\hat{\beta}_1) \end{aligned} \quad (1.7)$$

Das bedeutet, mit 95% Wahrscheinlichkeit liegt das wahre β_0 und β_1 in dem Intervall

$$\begin{aligned} &[\hat{\beta}_0 - 2 \cdot SE(\hat{\beta}_0), \hat{\beta}_0 + 2 \cdot SE(\hat{\beta}_0)] \\ &[\hat{\beta}_1 - 2 \cdot SE(\hat{\beta}_1), \hat{\beta}_1 + 2 \cdot SE(\hat{\beta}_1)] \end{aligned} \quad (1.8)$$

1.1.2 Beurteilung der Anpassungsgüte des Modells

Nach Berechnung der Regressionsfunktion ist es von Interesse, in welchem Ausmaß diese Funktion nun tatsächlich die zugrundeliegende Stichprobe widerspiegelt. Die Qualität einer LR wird typischerweise anhand von zwei zusammenhängenden Größen beurteilt, nämlich der Varianz der Residuen und dem Bestimmtheitsmaß R^2 .

Varianz der Residuen

Die Varianz der Residuen (engl. *residual standard error*), abgekürzt RSE, ist eine Schätzung der Standardabweichung der Residuen ϵ_i : Je kleiner der Fehler, desto besser ist das geschätzte Modell. Die Formel zur Berechnung des RSE lautet:

$$RSE = \sqrt{\frac{1}{n-2}RSS} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1.9)$$

RSE wird als ein Maß für die mangelnde Anpassung des Modells (1.2) an die Daten betrachtet. Wenn die mithilfe des Modells erhaltenen Vorhersagen sehr nahe an den wahren Ergebniswerten liegen - das heißt, wenn $\hat{y}_i \approx y_i$ für $i = 1, \dots, n$ - dann wird (1.9) klein sein, und wir können schließen, dass das Modell die Daten sehr gut erklärt. Andererseits, wenn \hat{y}_i sehr stark von y_i abweicht, dann kann RSE ziemlich groß sein, was darauf hinweist, dass das Modell die Daten nicht gut widerspiegelt.

Bestimmtheitsmaß R^2

Ein weiteres Maß für die Anpassungsgüte des Modells ist das **Bestimmtheitsmaß R^2** . Zur Berechnung dieses Maßes definieren wir die Größen RSS und TSS . RSS , die *erklärte Quadratsumme*, ist unter (1.3) definiert und gibt die Streuung der Schätzwerte \hat{y}_i um den Mittelwert \hat{y}_i an. TSS ist die sogenannte *erklärende bzw. totale Quadratsumme* und gibt die Streuung der Messwerte y_i um den Mittelwert \bar{y}_i an:

$$TSS = \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (1.10)$$

Mit dem Verhältnis beider Größen wird R^2 definiert:

$$\begin{aligned} R^2 &= 1 - \frac{RSS}{TSS} \\ &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \\ &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \end{aligned} \quad (1.11)$$

R^2 setzt die erklärte Varianz ins Verhältnis zur Gesamtvarianz. Um so höher der Anteil der erklärten Varianz, um so besser ist die Erklärungsleistung des Modells, d. h. dessen Anpassung an die Daten der Stichprobe. Liegt der Wert nahe bei 1, so existiert ein linearer Zusammenhang zwischen X und Y . Nimmt das Maß den Minimalwert 0 an, so existiert kein linearer Zusammenhang zwischen X und Y .

1.2 Multiple lineare Regression

Bislang wurde die lineare Abhängigkeit zweier Variablen behandelt, doch viele praktische Anwendungen erfordern die Berücksichtigung von mehr als nur einer unabhängigen Variable. Soll nun also eine Zielgröße Y als lineare Funktion mehrerer Einflussgrößen X_1, X_2, \dots, X_n beschrieben werden, so kommt die MLR zur Anwendung. Dafür wird jeder unabhängigen Variable ein separates Regressionsgewicht zugewiesen. Bei p verschiedenen Variablen X_i mit $i = 1, \dots, p$ hat das multiple lineare Regressionsmodell die Form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (1.12)$$

Das Modell lässt sich in Matrixschreibweise formulieren: $y = X\beta + \epsilon$. Die dazugehörigen Vektoren bzw. Matrizen sind wie folgt definiert:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad x = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad (1.13)$$

1.2.1 Schätzung der Regressionskoeffizienten

Wie auch bei der ELR sind die Regressionskoeffizienten β_1, \dots, β_p unbekannt und müssen geschätzt werden. Mit den Schätzungen $\hat{\beta}_1, \dots, \hat{\beta}_p$ können wir Vorhersagen mit folgender Formel vornehmen:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p \quad (1.14)$$

Methode der kleinsten Quadrate nach Gauß

Auch hier verwenden wir zur Schätzung der Regressionskoeffizienten die Methode der kleinsten Quadrate nach Gauß und wollen die Residuenquadratsumme RSS minimie-

ren:

$$\begin{aligned}
 RSS &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \sum_{j=1}^p \hat{\beta}_j x_{ij})^2 \\
 &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_p x_{ip})^2
 \end{aligned} \tag{1.15}$$

Man kann diese Größe auch in Matrixschreibweise definieren:

$$RSS = (y - X\hat{\beta})^T (y - X\hat{\beta}) \tag{1.16}$$

Die Werte $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$, die (1.15) bzw. (1.16) minimieren, sind die Schätzungen der Regressionskoeffizienten des Modells. Im Gegensatz zu den einfachen linearen Regressions-schätzungen aus (1.4) & (1.5) haben die Schätzungen der multiplen Regressionskoeffizienten etwas komplizierte Formen, die sich am einfachsten mit Hilfe der Matrix-Algebra darstellen lassen. Die Regressionskoeffizienten, die die RSS minimieren und so nach der Methode der kleinsten Quadrate berechnet werden, lauten:

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{1.17}$$

1.2.2 Beurteilung der Anpassungsgüte des Modells

Die Größen RSE und R^2 werden auf die gleiche Weise zur Beurteilung der Anpassungsgüte des Modells wie bei der ELR berechnet und interpretiert. Für RSE gilt

$$RSE = \sqrt{\frac{1}{n-p-1} RSS} \tag{1.18}$$

und für R^2 gilt weiterhin:

$$\begin{aligned}
 R^2 &= 1 - \frac{RSS}{TSS} \\
 &= 1 - \frac{(y - \hat{y})^T (y - \hat{y})}{(y - \bar{y})^T (y - \bar{y})}
 \end{aligned} \tag{1.19}$$

1.3 Nachteile der Methode der kleinsten Quadrate

Im folgenden Kapitel werden einige Möglichkeiten diskutiert, wie die Regressionsmodelle verbessert werden können, indem die Methode der kleinsten Quadrate durch einige alternative Anpassungsverfahren ersetzt wird. Wie wir sehen werden, können alternative Anpassungsverfahren eine bessere Vorhersagegenauigkeit und Modellinterpretierbarkeit liefern.

Vorhersagegenauigkeit

Unter der Voraussetzung, dass die wahre Beziehung zwischen der Zielvariable Y und den Parametern X_i mit $i = 1, \dots, n$ annähernd linear ist, weisen die Schätzungen der kleinsten Quadrate eine geringe Verzerrung auf. Wenn $n \gg p$ ist, d. h. wenn n , die Anzahl der Beobachtungen, viel größer ist als p , die Anzahl der Variablen, dann haben die Schätzungen der kleinsten Quadrate in der Regel auch eine geringe Varianz und sind daher für Testbeobachtungen gut geeignet. Wenn jedoch n nicht viel größer als p ist, kann es zu einer großen Variabilität in der Anpassung der kleinsten Quadrate kommen, was zu einer Überanpassung und folglich zu schlechten Vorhersagen für zukünftige Beobachtungen führt, die nicht beim Modelltraining verwendet wurden. Und wenn $p > n$ ist, dann gibt es keine eindeutige Schätzung des kleinsten quadratischen Koeffizienten mehr: Die Varianz ist unendlich, sodass die Methode überhaupt nicht verwendet werden kann. Durch Einschränkung oder Verkleinerung der geschätzten Koeffizienten kann die Varianz oft erheblich reduziert werden, und zwar auf Kosten einer vernachlässigbaren Erhöhung der Verzerrung. Dies kann zu erheblichen Verbesserungen in der Genauigkeit führen, mit der wir die Beobachtungen vorhersagen können, die nicht im Modelltraining verwendet wurden.

Modellinterpretierbarkeit

Es kommt häufig vor, dass einige oder viele der Variablen, die z.B. in einem MLR verwendet werden, in Wirklichkeit nicht mit der Zielvariable Y assoziiert sind. Die Einbeziehung solcher irrelevanten Variablen führt zu unnötiger Komplexität im resultierenden Modell. Durch das Entfernen dieser Variablen, d. h. durch das Setzen der entsprechenden Koeffizientenschätzungen auf Null, können wir ein Modell erhalten, das einfacher zu interpretieren ist. Nun ist es äußerst unwahrscheinlich, dass die Methode der kleinsten Quadrate Koeffizientenschätzungen ergibt, die genau null sind. Im folgenden Kapitel sehen wir uns einige Ansätze zur automatischen Durchführung der Variablenselektion an.

2 Regularisierung

Die Regularisierung ist eine Methode zur Verbesserung der Vorhersagegenauigkeit und der Modellinterpretierbarkeit. Bei diesem Ansatz wird ein Modell mit allen unabhängigen Variablen geschätzt. Die geschätzten Regressionskoeffizienten werden jedoch relativ zu den Schätzungen der Methode der kleinsten Quadrate gegen Null geschrumpft. Diese **Schrumpfung** (auch als **Regularisierung** bekannt) hat den Effekt, dass die Varianz reduziert wird. Je nachdem, welche Art von Schrumpfung durchgeführt wird, können einige der Koeffizienten als genau Null geschätzt werden. Daher können Schrumpfungsmethoden auch eine Variablenselektion durchführen. Die zwei bekanntesten Methoden zur Regularisierung einer LR sind die Ridge und Lasso Regression.

2.1 Ridge Regression

Wir erinnern uns an die Methode der kleinsten Quadrate: Die Regressionskoeffizienten werden so gewählt, dass die Residuenquadratsumme minimiert wird:

$$RSS = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \quad (2.1)$$

Die Ridge Regression, auch als L2-Regularisierung bekannt, ist dieser Methode sehr ähnlich, nur dass die Koeffizienten durch Minimierung einer etwas anderen Größe geschätzt werden. Insbesondere sind die Ridge Regressionskoeffizienten $\hat{\beta}^R$ die Werte, die folgende Größe minimieren

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2, \quad (2.2)$$

wobei λ als *Tuningparameter* bezeichnet wird.

Wie bei der Methode der kleinsten Quadrate wählt die Ridge Regression die Regressionskoeffizienten, die das Maß RSS minimiert. Hinzu kommt noch ein sogenannter *Strafterm* $\sum_j \beta_j^2$. Dieser ist sehr klein für β_1, \dots, β_p nahe 0 und hat den Effekt, die Schätzungen für β_j gegen 0 zu schrumpfen. Der Tuninparameter λ reguliert den Einfluss dieser Summe auf die Schätzungen der Regressionskoeffizienten und bestimmt somit die Höhe der Strafe. Wenn $\lambda = 0$ ist, so hat der Strafterm keine Auswirkung und die Ridge Regression liefert die Schätzungen der Methode der kleinsten Quadrate. Wenn jedoch $\lambda \rightarrow \infty$ ist, wächst die Auswirkung des Strafterms und die Schätzungen der Regressionskoeffizienten nähern sich Null. Im Gegensatz zur Methode der kleinsten Quadrate, die nur einen Satz von Koeffizientenschätzungen erzeugt, erzeugt die Ridge Regression für jeden Wert von λ einen anderen Satz von Koeffizientenschätzungen. Die Auswahl eines guten Wertes für λ ist entscheidend. Dies wird mittels der **Kreuzvalidierung** bestimmt (siehe Abschnitt 2.3).

2.2 Lasso Regression

Bei der Ridge Regression können die Regressionskoeffizienten zwar nahe 0 geschrumpft werden, aber können niemals ganz auf 0 gesetzt werden. Dies ist ein Nachteil für die Modellinterpretierbarkeit, da durch zu viele Variablen das Modell schnell unübersichtlich und kompliziert werden kann. Dieses Problem umgeht die Lasso Regression bzw. L1-Regularisierung. Die geschätzten Regressionskoeffizienten $\hat{\beta}_\lambda^L$ der Lasso Regression minimieren folgende Größe:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j| \quad (2.3)$$

Wie bei der Ridge Regression schrumpft das Lasso die Koeffizientenschätzungen gegen Null. Im Fall des Lasso hat der Strafterm jedoch den Effekt, dass einige der Koeffizientenschätzungen gezwungen werden, genau gleich Null zu sein, wenn der Tuningparameter λ ausreichend groß ist. Daher führt das Lasso eine Variablenauswahl durch. Infolgedessen sind die mit dem Lasso generierten Modelle im Allgemeinen viel einfacher zu interpretieren als die mit der Ridge Regression erzeugten, da eine geringere Anzahl von Variablen in das Modell einfließen. Wie bei der Ridge Regression ist die Wahl des besten Parameters λ entscheidend.

2.3 Kreuzvalidierung

Beide Regularisierungsverfahren erfordern eine Methode zur Auswahl eines Wertes für den Tuningparameter λ aus (2.2) und (2.3). Die Kreuzvalidierung, (engl. *cross validation*) ist solch eine Methode. Wir wählen ein Raster von λ -Werten und berechnen den sogenannten Kreuzvalidierungsfehler für jeden Wert von λ . Dann wird der Wert des Tuningparameters gewählt, für den der Kreuzvalidierungsfehler am kleinsten ist. Schließlich wird das Modell mit allen verfügbaren Beobachtungen und dem ausgewählten λ neu angepasst. Der Kreuzvalidierungsfehler ist definiert als

$$CV = \frac{1}{k} \sum_{i=1}^k (y_i - \hat{y}_i)^2 \quad (2.4)$$

Literatur

- [1] HASTIE, T. ; TIBSHIRANI, R. ; FRIEDMAN, J. : *The elements of statistical learning: data mining, inference and prediction.* 2. Springer
<https://web.stanford.edu/hastie/ElemStatLearn/>
- [2] JAMES, G. ; WITTEN, D. ; HASTIE, T. ; TIBSHIRANI, R. : *An Introduction to Statistical Learning: with Applications in R.* Springer
<https://faculty.marshall.usc.edu/gareth-james/ISL/>
- [3] WEISBERG, S. : *Applied linear regression.* Bd. 528. John Wiley & Sons, 2005

Eidesstattliche Versicherung

Name, Vorname: Demirel, Aylin

Hiermit erkläre ich an Eides statt, dass die vorliegende, an diese Erklärung angefügte Seminararbeit selbständig und ohne jede unerlaubte Hilfe angefertigt wurde, dass sie noch keiner anderen Stelle zur Prüfung vorgelegen hat und dass sie weder ganz noch im Auszug veröffentlicht worden ist. Die Stellen der Arbeit – einschließlich Tabellen, Karten, Abbildungen etc. – die anderen Werken und Quellen (auch Internetquellen) dem Wortlaut oder dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall als Entlehnung mit exakter Quellenangabe kenntlich gemacht.

17. 12. 2020

Datum

A. Demirel

Unterschrift