week3: Die Methode der kleinsten Quadrate bei p Regressoren: Projektion auf p-dimensionale Unterräume, Teil1

Gegeben seien die Datenvektoren

$$\vec{x}_1 = \begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{n,1} \end{pmatrix}, \quad \vec{x}_2 = \begin{pmatrix} x_{1,2} \\ x_{2,2} \\ \vdots \\ x_{n,2} \end{pmatrix}, \quad \cdots \quad, \vec{x}_p = \begin{pmatrix} x_{1,p} \\ x_{2,p} \\ \vdots \\ x_{n,p} \end{pmatrix}$$

und

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Wir möchten das $\vec{y}$ möglichst gut durch eine Linearkombination der $\vec{x}_1, \cdots, \vec{x}_p$ approximieren:

$$\vec{y} = \beta_1 \vec{x}_1 + \beta_2 \vec{x}_2 + \cdots + \beta_p \vec{x}_p + \text{error} \tag{1}$$

oder in Koordinaten

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \text{error}_i \tag{2}$$

Wenn wir noch einen konstanten Term in Gleichung (2) erlauben wollen, schreiben wir anstatt Gleichung (2) das folgende:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \text{error}_i \tag{3}$$

Um eine kompakte Notation in vektorschreibweise zu ermöglichen, definieren wir den Vektor

$$\vec{x}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

so dass wir anstatt von Gleichung (3) auch folgendes schreiben können:

$$\vec{y} = \beta_0 \vec{x}_0 + \beta_1 \vec{x}_1 + \beta_2 \vec{x}_2 + \cdots + \beta_p \vec{x}_p + \text{error} \tag{4}$$

Unter 'möglichst gut' approximieren wollen wir von jetzt an das Minimieren der L^2-Norm verstehen. Wir wollen die Regressionskoeffizienten $\beta_0, \beta_1, \cdots, \beta_p$ also durch die Forderung

$$F(\beta_0, \beta_1, \cdots, \beta_p) \;:=\; \left\| \vec{y} - (\beta_0 \vec{x}_0 + \beta_1 \vec{x}_1 + \cdots + \beta_p \vec{x}_p) \right\|^2 \;\stackrel{!}{\to}\; \min \tag{5}$$

bestimmen. In Koordinaten liest sich das dann so:

$$F(\beta_0, \beta_1, \cdots, \beta_p) \;=\; \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}) \right]^2 \;\stackrel{!}{\to}\; \min$$

oder, um die Notation etwas einheitlicher zu machen,

$$F(\beta_0, \beta_1, \cdots, \beta_p) \;=\; \sum_{i=1}^{n} \left[y_i - (\beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}) \right]^2 \;\stackrel{!}{\to}\; \min \tag{6}$$

da wir ja $x_{i,0} = 1$ definiert hatten. Notwendige Bedingung für ein Minimum ist

$$\nabla F \;=\; \left(\tfrac{\partial F}{\partial \beta_0}, \tfrac{\partial F}{\partial \beta_1}, \cdots, \tfrac{\partial F}{\partial \beta_p} \right) \;\stackrel{!}{=}\; 0 \tag{7}$$

Wir haben für $j \in \{0, 1, \cdots, p\}$:

$$\tfrac{\partial F}{\partial \beta_j} \;=\; \sum_{i=1}^{n} (-2 x_{i,j}) \left[y_i - (\beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}) \right] \;\stackrel{!}{=}\; 0 \tag{8}$$

Wegen

$$\sum_{i=1}^{n} x_{i,j}\, y_i \;=\; \vec{x}_j\, \vec{y}$$

ist Gleichung (8) äquivalent zu

$$\vec{x}_j \vec{y} - \left(\beta_0\, \vec{x}_j \vec{x}_0 + \beta_1\, \vec{x}_j \vec{x}_1 + \cdots + \beta_p\, \vec{x}_j \vec{x}_p \right) \;\stackrel{!}{=}\; 0 \qquad \forall j \in \{0, 1, \cdots, p\}$$

oder

$$\beta_0\, \vec{x}_j \vec{x}_0 + \beta_1\, \vec{x}_j \vec{x}_1 + \cdots + \beta_p\, \vec{x}_j \vec{x}_p \;=\; \vec{x}_j \vec{y} \qquad \forall j \in \{0, 1, \cdots, p\}$$

Damit erhalten wir das System

$$\begin{array}{rcl} \beta_0\, \vec{x}_0 \vec{x}_0 + \beta_1\, \vec{x}_0 \vec{x}_1 + \cdots + \beta_p\, \vec{x}_0 \vec{x}_p & = & \vec{x}_0 \vec{y} \\ \beta_0\, \vec{x}_1 \vec{x}_0 + \beta_1\, \vec{x}_1 \vec{x}_1 + \cdots + \beta_p\, \vec{x}_1 \vec{x}_p & = & \vec{x}_1 \vec{y} \\ & \vdots & \\ \beta_0\, \vec{x}_p \vec{x}_0 + \beta_1\, \vec{x}_p \vec{x}_1 + \cdots + \beta_p\, \vec{x}_p \vec{x}_p & = & \vec{x}_p \vec{y} \end{array} \tag{9}$$

oder in Matrix-Notation

$$A\, \vec{\beta} \;=\; \vec{b} \tag{10}$$

mit

$$A = \begin{pmatrix} \vec{x}_0\vec{x}_0 & \vec{x}_0\vec{x}_1 & \cdots & \vec{x}_0\vec{x}_p \\ \vec{x}_1\vec{x}_0 & \vec{x}_1\vec{x}_1 & \cdots & \vec{x}_1\vec{x}_p \\ \vdots & \vdots & & \vdots \\ \vec{x}_p\vec{x}_0 & \vec{x}_p\vec{x}_1 & \cdots & \vec{x}_p\vec{x}_p \end{pmatrix}, \qquad \vec{b} = \begin{pmatrix} \vec{x}_0\vec{y} \\ \vec{x}_1\vec{y} \\ \vdots \\ \vec{x}_p\vec{y} \end{pmatrix} \tag{11}$$

und dem Vektor der Regressionskoeffizienten

$$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \tag{12}$$

Die beta's sind damit gegeben durch

$$\vec{\beta} = A^{-1}\vec{b} \tag{13}$$

In dem folgendem Lemma halten wir fest, dass, wenn die Regressoren $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p$ linear unabhängig sind, die Matrix A auch tatsächlich invertierbar ist.

Lemma 3.1: Die Matrix

$$A = \begin{pmatrix} \vec{x}_0\vec{x}_0 & \vec{x}_0\vec{x}_1 & \cdots & \vec{x}_0\vec{x}_p \\ \vec{x}_1\vec{x}_0 & \vec{x}_1\vec{x}_1 & \cdots & \vec{x}_1\vec{x}_p \\ \vdots & \vdots & & \vdots \\ \vec{x}_p\vec{x}_0 & \vec{x}_p\vec{x}_1 & \cdots & \vec{x}_p\vec{x}_p \end{pmatrix}$$

ist genau dann invertierbar, wenn die Vektoren

$$\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p$$

linear unabhängig sind.

Beweis: Übungsblatt 3. ■

Im folgenden wollen wir immer annehmen, dass die Regressoren $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p$ linear unabhängig sind, und wir führen die folgende Notation ein: die Matrix X der Regressoren ist gegeben durch

$$X := \begin{pmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n\times(p+1)} \tag{14}$$

Damit können wir dann schreiben

$$\begin{aligned} X^TX &= \begin{pmatrix} - & \vec{x}_0 & - \\ - & \vec{x}_1 & - \\ & \vdots & \\ - & \vec{x}_p & - \end{pmatrix} \begin{pmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{pmatrix} \\ &= \begin{pmatrix} \vec{x}_0\vec{x}_0 & \vec{x}_0\vec{x}_1 & \cdots & \vec{x}_0\vec{x}_p \\ \vec{x}_1\vec{x}_0 & \vec{x}_1\vec{x}_1 & \cdots & \vec{x}_1\vec{x}_p \\ \vdots & \vdots & & \vdots \\ \vec{x}_p\vec{x}_0 & \vec{x}_p\vec{x}_1 & \cdots & \vec{x}_p\vec{x}_p \end{pmatrix} = A \end{aligned}$$

und

$$X^T\vec{y} \;=\; \begin{pmatrix} - & \vec{x}_0 & - \\ - & \vec{x}_1 & - \\ & \vdots & \\ - & \vec{x}_p & - \end{pmatrix} \begin{pmatrix} | \\ \vec{y} \\ | \end{pmatrix} \;=\; \begin{pmatrix} \vec{x}_0\vec{y} \\ \vec{x}_1\vec{y} \\ \vdots \\ \vec{x}_p\vec{y} \end{pmatrix} \;=\; \vec{b}$$

und bekommen

$$\vec{\beta} \;=\; A^{-1}\vec{b} \;=\; (X^TX)^{-1}X^T\vec{y} \tag{15}$$

Der Regression-Fit, das ist die rechte Seite von Gleichung (4) ohne den Error-Term, mit den berechneten beta's, ist dann gegeben durch

$$\begin{aligned} \text{RegressionFit} \;&:=\; \beta_0\,\vec{x}_0 \;+\; \beta_1\,\vec{x}_1 \;+\; \beta_2\,\vec{x}_2 \;+\; \cdots \;+\; \beta_p\,\vec{x}_p \\ &=\; \begin{pmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \\ &=\; X\vec{\beta} \\ &=\; X(X^TX)^{-1}X^T\vec{y} \qquad (16) \\ &=:\; P_X\vec{y} \end{aligned}$$

mit der $n \times n$ Matrix

$$P_X \;=\; X(X^TX)^{-1}X^T \tag{17}$$

Fassen wir unsere Resultate in einem Theorem zusammen:

Theorem 3.2: Gegeben seien die Datenvektoren $\vec{y} \in \mathbb{R}^n$ und $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p \in \mathbb{R}^n$ mit $n \geq p+1$ wobei die Regressoren $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p$ linear unabhängig seien, die Matrix der Regressoren

$$X \;:=\; \begin{pmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{pmatrix} \;\in\; \mathbb{R}^{n\times(p+1)}$$

habe also maximalen Rang. Wir betrachten das Minimierungsproblem

$$\begin{aligned} F(\vec{\beta}) \;=\; F(\beta_0, \beta_1, \cdots, \beta_p) \;&=\; \big\|\, \vec{y} \;-\; (\beta_0\vec{x}_0 \;+\; \beta_1\vec{x}_1 \;+\; \cdots \;+\; \beta_p\vec{x}_p)\,\big\|^2 \\ &=\; \big\|\, \vec{y} \;-\; X\vec{\beta}\,\big\|^2 \;\stackrel{!}{\to}\; \min \end{aligned}$$

Dann gilt: Das F wird durch die folgende Wahl der beta's minimiert,

$$\vec{\beta}_{\min} \;=\; (X^TX)^{-1}X^T\vec{y}$$

und die beste Approximation an das $\vec{y}$, der Regression-Fit $\vec{y}_{\min} := X\vec{\beta}_{\min}$, ist gegeben durch

$$\vec{y}_{\min} \quad = \quad P_X \vec{y}$$

mit der linearen Abbildung

$$\begin{aligned} P_X &\quad : \quad \mathbb{R}^n \to \mathbb{R}^n \,, \\ P_X &\quad = \quad X(X^T X)^{-1} X^T \,. \end{aligned}$$

In der nächsten Vorlesung werden wir uns dann das P_X etwas genauer anschauen und eine geometrische Interpretation angeben.