week4: Die Methode der kleinsten Quadrate bei p Regressoren, Teil2

Letztes Mal hatten wir das folgende Theorem bewiesen:

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 $p+1$ Regressoren $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p$ linear unabhängig seien, die Matrix der Regressoren

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

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

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

mit der linearen Abbildung

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

Betrachten wir den $p+1$ dimensionalen linearen Unterraum im $\mathbb{R}^n$, der durch die $p+1$ Regressoren $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p$ aufgespannt wird,

$$L_X \;:=\; \left\{ \sum_{j=0}^{p} \lambda_j\vec{x}_j \;\middle|\; \lambda_0, \lambda_1, \cdots, \lambda_p \in \mathbb{R} \right\} \;\subset\; \mathbb{R}^n \tag{2}$$

Dann gilt das folgende

Theorem 4.1 (Geometrische Interpretation): Es sei X die Matrix der Regressoren gegeben durch (1) und L_X der von den Regressoren aufgespannte lineare Unterraum im $\mathbb{R}^n$. Weiter sei

$$\begin{aligned} P_X &: \mathbb{R}^n \to \mathbb{R}^n , \\ P_X &= X(X^TX)^{-1}X^T . \end{aligned} \tag{3}$$

Dann gilt: Das P_X ist die orthogonale Projektion auf L_X und

$$P_{X^\perp} := Id - P_X \tag{4}$$

ist die orthogonale Projektion auf das orthogonale Komplement von L_X. Genauer gelten die folgenden Eigenschaften:

a) $P_X\vec{x}_j = \vec{x}_j \quad \forall j = 0, \cdots, p$

b) $P_X^2 = P_X$

c) $P_X^T = P_X$

d) $P_X P_{X^\perp} = P_{X^\perp} P_X = 0$

Beweis: a) Wir haben

$$X^T\vec{x}_j = \begin{pmatrix} - & \vec{x}_0 & - \\ - & \vec{x}_1 & - \\ & \vdots & \\ - & \vec{x}_p & - \end{pmatrix} \begin{pmatrix} | \\ \vec{x}_j \\ | \end{pmatrix} = \begin{pmatrix} \vec{x}_0\vec{x}_j \\ \vec{x}_1\vec{x}_j \\ \vdots \\ \vec{x}_p\vec{x}_j \end{pmatrix} = \big(X^TX \big)_{\text{j−te Spalte}}$$

Damit bekommen wir, wenn Id_{p+1} die $p+1$ dimensionale Einheitsmatrix bezeichnet,

$$\begin{aligned} (X^TX)^{-1}X^T\vec{x}_j &= (X^TX)^{-1}\big(X^TX \big)_{\text{j−te Spalte}} \\ &= \big(Id_{p+1} \big)_{\text{j−te Spalte}} \\ &= \big(\delta_{i,j} \big)_{i=0}^{p} \end{aligned}$$

und damit

$$\begin{aligned} P_X\vec{x}_j &= X(X^TX)^{-1}X^T\vec{x}_j \\ &= X\big(\delta_{i,j} \big)_{i=0}^{p} \\ &= \begin{pmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{pmatrix} \big(\delta_{i,j} \big)_{i=0}^{p} = \begin{pmatrix} | \\ \vec{x}_j \\ | \end{pmatrix} = \vec{x}_j \end{aligned}$$

b) Wir haben

$$\begin{aligned} P_X^2 &= X(X^TX)^{-1}X^T \;\; X(X^TX)^{-1}X^T \\ &= X(X^TX)^{-1}\underbrace{X^TX(X^TX)^{-1}}_{=Id}\,X^T \\ &= X(X^TX)^{-1}X^T \;=\; P_X \end{aligned}$$

c) Es gilt

$$\begin{aligned} P_X^T &= \left[\,X(X^TX)^{-1}X^T\,\right]^T \\ &= (X^T)^T\,\left[(X^TX)^{-1}\right]^T\,X^T \\ &= X\,\left[(X^TX)^T\right]^{-1}X^T \\ &= X\,\left[\,X^TX\,\right]^{-1}X^T \;=\; P_X \end{aligned}$$

d) Schliesslich, mit $P_X^2 = P_X$,

$$P_{X^\perp}P_X \;=\; (Id - P_X)P_X \;=\; P_X - P_X^2 \;=\; P_X - P_X \;=\; 0$$

und $P_X P_{X^\perp} = 0$ mit einer analogen Rechnung. Damit können wir für jedes $\vec{y} \in \mathbb{R}^n$ schreiben

$$\begin{aligned} \vec{y} &= Id\,\vec{y} \\ &= P_X\,\vec{y} \;+\; (Id - P_X)\,\vec{y} \\ &=: \vec{y}_1 \;+\; \vec{y}_2 \end{aligned}$$

mit $\vec{y}_1 \cdot \vec{y}_2 \;=\; 0$ und $\vec{y}_1 \in L_X$, $\vec{y}_2 \in L_{X^\perp}$. Damit ist das Theorem bewiesen. $\blacksquare$

Beispiel: Schauen wir uns noch ein kleines Beispiel an: Gegeben seien die Vektoren oder Regressoren

$$\vec{x}_1 \;:=\; \begin{pmatrix} 1 \\ a \\ 0 \end{pmatrix}, \qquad \vec{x}_2 \;:=\; \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

mit einem $a \neq 1$ und

$$\vec{y} \;:=\; \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Wir betrachten das Regressionsproblem

$$\vec{y} \;=\; \beta_1\,\vec{x}_1 \;+\; \beta_2\,\vec{x}_2 \;+\; error/noise$$

Bestimmen Sie die Regressionskoeffizienten β_1 und β_2 sowie die Bestapproximation

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

an das $\vec{y}$. Bestimmen Sie ebenfalls das P_X und das $P_{X^\perp}$.

Lösung: Machen wir an der Tafel.