VL6: Die Methode der kleinsten Quadrate bei p Regressoren, Teil2 und Anwendungsbeispiel: Fourierreihenentwicklung

Letztes Mal hatten wir das folgende Theorem bewiesen:

Theorem 5.2: 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 6.1 (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

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

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

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

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

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

und damit

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

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 $P_XP_{X^\perp} = 0$ 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. ■

Anwendungsbeispiel: Fourierreihenentwicklung

Erinnerung Fourierreihen: Jede integrierbare Funktion[1]

$$f\ :\ [-\pi, +\pi]\ \ \to\ \ \mathbb{R}$$

kann man in eine Fourierreihe entwickeln:

$$f(x) \ \ = \ \ \tfrac{a_0}{2}\ +\ \sum_{k=1}^{\infty}\left\{\,a_k\,\cos kx\ +\ b_k\,\sin kx\,\right\} \tag{5}$$

mit den Entwicklungskoeffizienten

$$a_k \ \ = \ \ \tfrac{1}{\pi}\textstyle\int_{-\pi}^{\pi} f(x)\,\cos kx\,dx\ , \qquad\qquad k = 0, 1, 2, \cdots \tag{6}$$

$$b_k \ \ = \ \ \tfrac{1}{\pi}\textstyle\int_{-\pi}^{\pi} f(x)\,\sin kx\,dx\ , \qquad\qquad k = 1, 2, \cdots \tag{7}$$

[1]das kann auch ein anderes Intervall sein, wir nehmen der Einfachheit halber $[-\pi, \pi]$

Dabei bilden die Funktionen

$$\begin{array}{ccccccccc} 1 & , & \cos x & , & \cos 2x & , & \cos 3x & , & \cdots \\ & & \sin x & , & \sin 2x & , & \sin 3x & , & \cdots \end{array}$$

eine Orthogonalbasis von $L^2\big([-\pi,\pi]\big)$, das ist die Menge der quadrat-integrablen Funktionen auf dem Intervall $[-\pi,\pi]$. Genauer gelten die folgenden Orthogonalitätsrelationen, die folgenden Integrale:

$$\begin{aligned} \int_{-\pi}^{\pi} \cos \ell x \, \cos mx \, dx &= \pi \, \delta_{\ell,m} \\ \int_{-\pi}^{\pi} \sin \ell x \, \sin mx \, dx &= \pi \, \delta_{\ell,m} \end{aligned}$$

und

$$\int_{-\pi}^{\pi} \cos \ell x \, \sin mx \, dx = 0$$

mit natürlichen Zahlen ℓ und m.

<u>Formulierung als Regressionsproblem</u>: Es sei jetzt

$$f \; : \; [-\pi, +\pi] \;\to\; \mathbb{R}$$

eine gegebene Funktion. Wir diskretisieren das Intervall $[-\pi,\pi]$ mit Schrittweite Δx:

$$-\pi =: x_0,\; x_1 = x_0 + \Delta x,\; x_2 = x_0 + 2\Delta x,\; \cdots\; x_n = x_0 + n\Delta x = +\pi$$

mit

$$\Delta x = \tfrac{+\pi-(-\pi)}{n} = \tfrac{2\pi}{n}$$

Unser Datenvektor $\vec{y}$ ist dann

$$\vec{y} = \begin{pmatrix} f(x_0) \\ \vdots \\ f(x_n) \end{pmatrix} \in \mathbb{R}^{n+1}$$

und die Regressoren sind gegeben durch

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} \cos x_0 \\ \vdots \\ \cos x_n \end{pmatrix}, \begin{pmatrix} \cos 2x_0 \\ \vdots \\ \cos 2x_n \end{pmatrix}, \;\cdots\;, \begin{pmatrix} \cos px_0 \\ \vdots \\ \cos px_n \end{pmatrix}$$

und

$$\begin{pmatrix} \sin x_0 \\ \vdots \\ \sin x_n \end{pmatrix}, \begin{pmatrix} \sin 2x_0 \\ \vdots \\ \sin 2x_n \end{pmatrix}, \;\cdots\;, \begin{pmatrix} \sin px_0 \\ \vdots \\ \sin px_n \end{pmatrix}$$

Definieren wir etwa die Matrizen

$$X_{\cos} \;:=\; \begin{pmatrix} | & | & & | \\ \cos x_i & \cos 2x_i & \cdots & \cos px_i \\ | & | & & | \end{pmatrix} \;\in\; \mathbb{R}^{(n+1)\times p}$$

$$X_{\sin} \;:=\; \begin{pmatrix} | & | & & | \\ \sin x_i & \sin 2x_i & \cdots & \sin px_i \\ | & | & & | \end{pmatrix} \;\in\; \mathbb{R}^{(n+1)\times p}$$

dann ist die Fourierreihenentwicklung (5), wenn wir sie nach p Termen abbrechen,

$$f(x) \;=\; \tfrac{a_0}{2} \;+\; \sum_{k=1}^{p} \big\{\, a_k \, \cos kx \;+\; b_k \, \sin kx \,\big\} \;+\; \text{error} \tag{8}$$

äquivalent zu

$$\begin{pmatrix} | \\ f(x_i) \\ | \end{pmatrix} \;=\; \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \frac{a_0}{2} \;+\; X_{\cos} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \;+\; X_{\sin} \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix} \;+\; \text{error} \tag{9}$$

$$\Leftrightarrow \qquad \vec{y} \;=\; X\vec{\beta} \;+\; \text{error}$$

mit der Matrix X der Regressoren gegeben durch, mit $\vec{x}_0 = (1, 1, \cdots, 1) \in \mathbb{R}^{n+1}$,

$$X \;:=\; \begin{pmatrix} \vec{x}_0 & X_{\cos} & X_{\sin} \end{pmatrix} \;\in\; \mathbb{R}^{(n+1)\times(2p+1)}$$

und den Regressionskoeffizienten

$$\vec{\beta} \;:=\; \big(a_0/2, a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_p \big) \;\in\; \mathbb{R}^{2p+1}$$

Mit Hilfe des `lm()`-Befehls aus der R-Software lassen sich die Regressionskoeffizienten dann sehr einfach bestimmen.

$\rightarrow$ Start R-Session.

Schauen wir uns noch an, wie die allgemeine Formel für die beta's in diesem Fall konkret aussieht: Wir hatten

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

mit

$$X^T X \;=\; \big(\, \vec{x}_\ell \cdot \vec{x}_m \,\big)$$

gegeben durch die Skalarprodukte der Regressoren. Betrachten wir der Einfachheit halber etwa eine reine Sinus-Entwicklung so dass $X = X_{\sin}$. Dann haben wir etwa, wenn $X_{\sin,\ell}$ die ℓ'te Spalte der Regressionsmatrix bezeichnet,

$$\begin{aligned} X_{\sin,\ell} \cdot X_{\sin,m} &= \sum_{i=0}^{n} \sin \ell x_i \, \sin m x_i \\ &= \frac{1}{\Delta x} \, \Delta x \, \sum_{i=0}^{n} \sin \ell x_i \, \sin m x_i \\ &\approx \frac{1}{\Delta x} \, \int_{-\pi}^{\pi} \sin \ell x \, \sin m x \, dx \\ &= \frac{\pi}{\Delta x} \, \delta_{\ell,m} \end{aligned}$$

Also haben wir, bei einer reinen Sinus-Entwicklung,

$$X^T X \approx \tfrac{\pi}{\Delta x} \, Id$$

und damit

$$(X^T X)^{-1} \approx \tfrac{\Delta x}{\pi} \, Id$$

Weiterhin ist

$$X^T \vec{y} = \begin{pmatrix} | \\ \vec{x}_\ell \cdot \vec{y} \\ | \end{pmatrix}$$

mit

$$\begin{aligned} \vec{x}_\ell \cdot \vec{y} &= X_{\sin,\ell} \cdot \begin{pmatrix} | \\ f(x_i) \\ | \end{pmatrix} = \sum_{i=0}^{n} \sin(\ell x_i) \, f(x_i) \\ &\approx \frac{1}{\Delta x} \, \int_{-\pi}^{\pi} \sin(\ell x) \, f(x) \, dx \\ &\overset{(7)}{=} \frac{\pi}{\Delta x} \, b_\ell \end{aligned}$$

Also liefert die allgemeine Formel für die Regressionskoeffizienten in diesem konkreten Fall

$$\begin{aligned} \vec{\beta} &= (X^T X)^{-1} X^T \vec{y} \\ &\approx \frac{\Delta x}{\pi} \, Id \, \frac{\pi}{\Delta x} \, \begin{pmatrix} | \\ b_\ell \\ | \end{pmatrix} \\ &= \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix} \end{aligned}$$

und wir erhalten genau die Entwicklungskoeffizienten der Fourierreihenentwicklung.