PDV - 9.10.2026

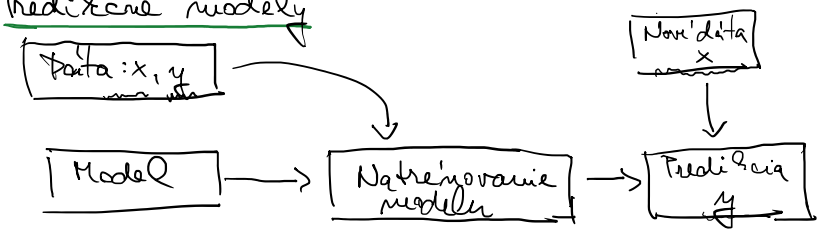
piatok 9. októbra 2026 10:33

k – medoids, $\forall j$, $k^* = \arg\min \sum_{i:c_i=j} \|x_i - x_k\| \Rightarrow \mu_j = x_{k^*}$

Nevýhoda: zhlukovanie

- získané zhlukovanie nemusí byť "správne"

Predikčné modely

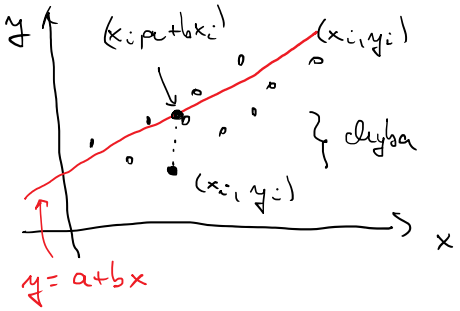


Lineárna regresia

1805 – Legendre, 1809 – Gauss, meno: Galton (1822 – 1911) ⟨ Ekonometria ⟩

- regresný problém: $y_i \in \mathbb{R}$

priamková regresia: napr.: výška → váha



priamka: $y \approx a + bx$

$a, b = ?$

→ metóda najmenších štvorcov:

$$\sum_{i=1}^{n} (y_i - a - bx_i)^2 \longrightarrow \min_{a,b}$$

$J(a,b)$

- aby priamka bola čo najbližšie k dátam

Viac premenných:

váha$_i \approx a + b_1$ veľk$_i + b_2$ výška$_i + \ldots + b_6 \cdot$ pár$_i$

všeob.: $y_i \approx \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_{p-1} x_{i,p-1}$

($x_{i,1}$: veľk$_i$; $x_{i,p-1}$: pár$_i$)

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix} = ?$$

$$\Rightarrow J(\beta) = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_{p-1} x_{i,p-1}) \right]^2 \longrightarrow \min_{\beta}$$

skut. y_i; odhadovaná y_i modelom; chyba$_i$

Otázky:

1.) – aký je vzťah medzi výškou a váhou?
– aký je silný?
– viem predikovať váhu na základe výšky?

2.) – viem predikovať váhu na základe ostatných premenných?
– aký je vplyv ktorej premennej?

β_0: priesečník s y-ovou osou ⇒ "intercept"

zistili sme $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -68 \\ 0{,}9 \end{pmatrix}$ ⇒ váha$_i \approx -68 + 0{,}9 \cdot$ výška$_i$

nárast výšky o 1 cm spôsobí nárast

$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \approx \begin{pmatrix} 0,9 \end{pmatrix}$

nárast výšky o 1 cm spôsobí nárast odkladovanej váhy o 0,9 kg

$\beta = \begin{pmatrix} -117 \\ -0,1 \\ 0,5 \\ \vdots \\ 0,7 \end{pmatrix} \Rightarrow$ váha$_i \approx -117 - 0,1 \cdot$ vek$_i + 0,5 \cdot$ výška$_i + \ldots + 0,7$ pás$_i$

nárast k veku o 1 rok spôsobí pokles odkladovanej váhy o $0,1$ kg (ak ostatné premenné sa nemenia)

vektorový zápis:

1.) priamková: $y_1 \approx a + bx_1$ ... $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \approx \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$

$y_n \approx a + bx_n$ $\qquad y \approx X\beta$

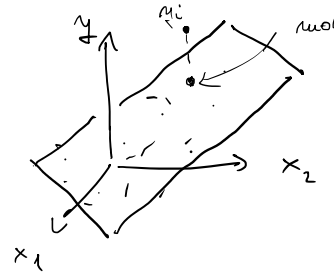
2.) všeob.: $y_1 \approx \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_{p-1} x_{1,p-1}$

$y_n \approx \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \ldots + \beta_{p-1} x_{n,p-1}$

$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \approx \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ \vdots & & & & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$

(vektory výšky; y, X, β)

$y \approx X\beta$



modelované y_i: $\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2}$

$\Rightarrow$ lin. reg.: $y \approx X\beta$

[štatistici: $y = X\beta + \varepsilon$, ε chyby]

1.) $J(a,b) = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$

vektor chýb $= \begin{pmatrix} y_1 - (a+bx_1) \\ \vdots \\ y_n - (a+bx_n) \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} a+bx_1 \\ \vdots \\ a+bx_n \end{pmatrix} = y - X\beta$

$\Rightarrow J(a,b) = \sum_{i=1}^{n} ($ prvky vektora chýb $)^2 = \sum_{i=1}^{n} ((y - X\beta)_i)^2 = \|y - X\beta\|^2$

$\left[\|z\|^2 = z^T z = \sum z_i^2\right]$

2.) $J(\beta) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i + \ldots))^2 = \sum_{i=1}^{n} (\text{chyba}_i)^2 = \sum_{i=1}^{n} ((y - X\beta)_i)^2 = \|y - X\beta\|^2 \rightarrow \min_{\beta}$

Riešenie: $\beta = (X^T X)^{-1} X^T y$ ... riešenie normálnych rovníc $(X^T X)\beta = X^T y$

geometricky:

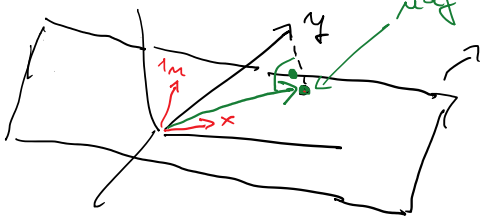
1.) $\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \end{bmatrix} a + \begin{bmatrix} x_1 \\ \vdots \end{bmatrix} b$

$y \approx \ldots$

$$y \approx \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} a + \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} b$$

1_n ... "po stĺpcoch" ... x

$\mathbb{R}^n$:

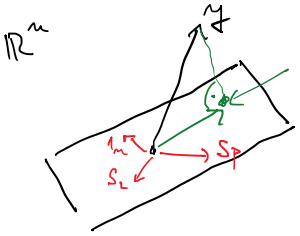


najlepšia $X\beta$: kolmá projekcia y do $\mathcal{C}(X)$

všetky $a \cdot 1_n + b \cdot x = \text{všetky } X\beta = \mathcal{C}(X)$: stĺpcový priestor X

$\Rightarrow$ normálne rovnice $X^T X\beta = X^T y$

podobne pre všeobecné $X = [1_n \; s_1 \; \dots \; s_p]$



optimálne $X\beta$

$\mathcal{C}(X)$: "p-rozmerná rovina"

predikcia: máme nové x $\rightarrow$ odhad y modelom?

$x = \begin{pmatrix} \text{veľ} \\ \text{výška} \\ \vdots \\ \text{pás} \end{pmatrix}$ $\Rightarrow$ predikovaná hodnota: $\hat{y} = \beta_0 + \beta_1 \cdot \text{veľ} + \dots + \beta_6 \text{pás}$

$= -118 - 0{,}1 \cdot \text{veľ} + \dots + 0{,}7 \text{pás}$

pozor: ak $x \cdot 12 \mapsto \beta \cdot \frac{1}{12}$

$\text{veľ} \cdot 12 \mapsto \beta_2 \cdot \frac{1}{12}$

$\Rightarrow$ interpretácia záleží od mierky

$\beta_{\text{veľ}} = -0{,}09$

$\beta_{\text{obv. ruky}} = 1{,}5$ $\Rightarrow$ obvod ruky má >10-krát väčší vplyv? $\Rightarrow$ Ani nie (Technicky áno) (Alebo záleží od mierky)

$\rightarrow$ Ako korektne porovnať vplyvy? — štandardizácia premenných: $\frac{\text{veľ} - \overline{\text{veľ}}}{\text{s.d.(veľ)}}$ apod.

"štandardizácia" β $\rightarrow$ neskôr

z-score