0 Hřebenová regrese

Hřebenová nebo také L_2 regularizace, stejně jako lineární regrese, předpokládá lineární model $Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon$. Nicméně navíc řeší problém kolinearity tím, že penalizuje vysoké hodnoty koeficientů $\boldsymbol{w}$ vyjma interceptu.

0.1 Regularizovaný reziduální součet čtverců

Hřebenová regrese tedy minimalizuje následují reziduální součet čtverců

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2$$

s regulačním členem $\lambda \geq 0$.

Pro $\lambda = 0$ dostáváme klasickou neregularizovanou lineární regresi. Intercept se nepenalizuje, protože ten jen zajišťuje $\mathrm{E}\,\varepsilon = 0$.

0.2 Minimalizace regularizovaného RSS

Pro účely hřebenové regrese zavedeme

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1}$$

Regularizovaný RSS_λ lze vyjádřit Jako

$$\begin{aligned} \mathrm{RSS}_\lambda(\boldsymbol{w}) &= \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2 \\ &= \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w} \end{aligned}$$

Nalezneme parciální derivaci a následně gradient

$$\begin{aligned} \frac{\partial\,\mathrm{RSS}(\boldsymbol{w})}{\partial w_j} &= \sum_{i=1}^{N} 2(Y_i - \boldsymbol{w}^T\boldsymbol{x}_i)(-x_{i,j}) + 2\lambda w_j \\ \nabla\,\mathrm{RSS}(\boldsymbol{w}) &= \sum_{i=1}^{N} 2(Y_i - \boldsymbol{w}^T\boldsymbol{x}_i)(-\boldsymbol{x}_i) + 2\lambda\mathbf{I}'\boldsymbol{w} \\ &= -2\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{w}) + 2\lambda\mathbf{I}'\boldsymbol{w} \end{aligned}$$

A položením rovno nule obdržíme ekvivalent normální rovnice

$$\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - \lambda\mathbf{I}'\boldsymbol{w} = 0$$
$$\boldsymbol{X}^T\boldsymbol{Y} - (\boldsymbol{X}^T\boldsymbol{X} + \lambda\mathbf{I}')\boldsymbol{w} = 0$$

Hessova matice zde vychází

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2(\boldsymbol{X}^T\boldsymbol{X} - \lambda\mathbf{I}')$$

Pro libovolné $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}, \lambda > 0$ platí

$$\begin{aligned}\boldsymbol{s}^T(\boldsymbol{X}^T\boldsymbol{X} + \lambda\mathbf{I}')\boldsymbol{s} &= \boldsymbol{s}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{s} + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s} \\ &= \|\boldsymbol{X}\boldsymbol{s}\|^2 + \lambda\sum_{i=1}^{p} s_i^2 > 0,\end{aligned}$$

Matice $\boldsymbol{X}^T\boldsymbol{X} + \lambda\mathbf{I}'$ je tedy vždy pozitivně definitní a regulární a řešení normální rovnice pro regulované RSS_λ je jednoznačné.

$$\hat{\boldsymbol{w}}_\lambda = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\mathbf{I}')^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

Predikce modelu je $\hat{Y} = \hat{\boldsymbol{w}}_\lambda^T\boldsymbol{x}$.

0.3 Modely bázových funkcí

Model lineární regrese, případně hřebenové regrese, modeluje pouze lineární funkci. Množinu příznaků lze však rozšířit jejími transformovanými variantami.

0.3.1 Bázové funkce

Pro $M \in \mathbb{N}$ zvolme M funkcí $\varphi_1, \ldots, \varphi_M$ z $\mathbb{R}^p$ do $\mathbb{R}$ reprezentující transformace. Těmto funkcím říkáme bázové funkce.

Modely bázových funkcí se od předchozích lineárních regresních modelů liší pouze tím, že místo původního $\boldsymbol{x}$ pracuje s $\boldsymbol{\varphi}(\boldsymbol{x}) := (1, \varphi_1(\boldsymbol{x}), \ldots, \varphi_M(\boldsymbol{x}))^T$

0.3.2 Volba bázových funkcí

Časté volby bázových funkcí jsou

- $\varphi(\boldsymbol{x}) = x_i$ – původní příznaky
- $\varphi(\boldsymbol{x}) = x_i^2, x_ix_j$ – mocniny, součiny (polynomiální regrese)
- $\varphi(\boldsymbol{x}) = \log(x_i), \sqrt{x_i}, \sin(x_i)$ – nelineární transformace
- $\varphi(\boldsymbol{x}) = \mathbb{1}_A$ – indikátory (dělení prostoru)

0.3.3 Model

Model bázové funkce rozšiřuje možnosti hřebenové regrese a může jej dělat velmi mocným a sofistikovaným nástrojem.

$$\boldsymbol{x} \longrightarrow \boldsymbol{\varphi}(\boldsymbol{x})\colon \quad Y = \boldsymbol{w}_\lambda^T \boldsymbol{\varphi}(\boldsymbol{x}) + \varepsilon$$

Postup s transformovaným vektorem $\boldsymbol{x}' := \boldsymbol{\varphi}(\boldsymbol{x})$ je dál zcela analogický jako u hřebenové regrese a predikce modelu je $\hat{Y} = \hat{\boldsymbol{w}}_\lambda^T \boldsymbol{\varphi}(\boldsymbol{x})$.