0 Statistické vlastnosti modelů

V modelu pro trénovací množinu $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ je $\boldsymbol{\varepsilon}$ náhodný vektor, z čehož plyne, že $\boldsymbol{Y}$ je náhodný vektor, a tedy i $\hat{\boldsymbol{w}}_\lambda$ je náhodný vektor.

0.1 Očekávaná chyba modelu

Protože je odhad $\hat{\boldsymbol{w}}_\lambda$ náhodný vektor, můžeme pomocí kvadratické ztrátové funkce zkoumat očekávanou chybu predikce $Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon$ pomocí $\hat{Y} = \hat{\boldsymbol{w}}_\lambda^T\boldsymbol{x}$.

0.1.1 Rozklad očekávané chyby modelu

Předpokládáme-li, nezávislost trénovací množiny s testovací množinou, tedy i nezávislost Y a $\hat{Y}$, pak lze očekávanou chybu spočíst jako

$$\mathrm{E}\, L(Y, \hat{Y}) = \mathrm{E}(Y - \hat{Y})^2 = \ldots = \operatorname{var} Y + \mathrm{E}(\hat{Y} - \mathrm{E}\, Y)^2 = \sigma^2 + \mathrm{MSE}(\hat{Y})$$

První člen odpovídá Bayesovské chybě, která je dána náhodností modelu. Druhý člen značíme $\mathrm{MSE}(\hat{Y})$ (mean squared error), který se dá dále dělit na dva členy:

$$\mathrm{MSE}(\hat{Y}) = \mathrm{E}(\hat{Y} - \mathrm{E}\, Y)^2 = \ldots = (\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y)^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\, \hat{Y})^2 = (\operatorname{bias} \hat{Y})^2 + \operatorname{var} \hat{Y}$$

kde $\operatorname{bias} \hat{Y}$ značí vychýlení odhadu $\hat{Y}$ a $\operatorname{var} \hat{Y}$ jeho rozptyl. Finální dekompozice očekávané chyby odhadu je

$$\mathrm{E}\, L(Y, \hat{Y}) = \sigma^2 + (\operatorname{bias} \hat{Y})^2 + \operatorname{var} \hat{Y}$$

a skládá se z neodstranitelné Bayesovské chyby. kvadrátu vychýlení odhadu a rozptylu odhadu.

Tento rozklad je pouze teoretický, v praxi máme jen celkové MSE, které se snažíme minimalizovat.

0.2 Bias-variance tradeoff

Zpravidla s komplexitou modelu klesá bias (vychýlení), ale roste variance (rozptyl) – model se přeučuje.

U hřebenové regrese s rostoucím regulačním koeficientem klesá rozptyl (regularizace je silnější, odhad $\hat{\boldsymbol{w}}$ je stabilnější), ale zároveň roste vychýlení (koeficienty více "dusíme" a jsou systematicky podhodnocované). Takovému chování v závislosti na hyperparametrech modelu se nazývá bias-variance tradeoff.

V praxi při ladění hřebenové regresi hledáme optimální λ, při které je MSE na validační množině $(Y_i', \boldsymbol{x}_i')$ nejmenší. MSE odhadujeme

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i' - \hat{\boldsymbol{w}}_\lambda^T \boldsymbol{x}_i')^2$$

Při použití hřebenové regrese je rozumné příznaky standardizovat, aby byly rozsahově podobné a penalizované stejně. Příznaky X_i nahradíme

$$X_i \leftarrow \frac{X_i - \bar{X}_i}{\sqrt{s^2_{X_i}}}, \quad \text{kde} \quad \bar{X}_i = \frac{1}{N}\sum_{i=1}^{N}(X_i)_j \quad \text{a} \quad s^2_{X_i} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X}_i)^2$$

0.3 Nestrannost odhadu v metodou nejmenších čtverců

Protože je $\hat{\boldsymbol{w}}_{\text{OLS}}$ bodový odhad (statistika) můžeme zkoumat zda je nestranný. Předpokládáme-li $\mathrm{E}\,\boldsymbol{\varepsilon} = 0$, pak z linearity střední hodnoty platí

$$\mathrm{E}\,\boldsymbol{Y} = \mathrm{E}(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{\varepsilon}) = \mathrm{E}\,\boldsymbol{X}\boldsymbol{w} + \mathrm{E}\,\boldsymbol{\varepsilon} = \boldsymbol{X}\boldsymbol{w}$$

Dále platí

$$\begin{aligned}
\mathrm{E}\,\hat{\boldsymbol{w}}_{\text{OLS}} &= \mathrm{E}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\,\mathrm{E}[\boldsymbol{Y}] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} = \boldsymbol{w}
\end{aligned}$$

Z čehož plyne

$$\mathrm{E}\,\hat{Y} = \mathrm{E}(\hat{\boldsymbol{w}}^T_{\text{OLS}}\boldsymbol{x}) = \mathrm{E}(\hat{\boldsymbol{w}}_{\text{OLS}})^T\boldsymbol{x} = \boldsymbol{w}^T\boldsymbol{x} = \mathrm{E}\,Y$$

$\hat{Y}$ je tedy nestranným bodovým odhadem EY. To znamená, že pro neregularizovanou lineární regresi platí, že je vychýlení nulové:

$$\operatorname{bias}\hat{Y} = \mathrm{E}\,\hat{Y} - \mathrm{E}\,Y = 0.$$