0 Logistická regrese

Logistická regrese predikuje pravděpodobnost jednotlivých hodnot vysvětlované proměnné. Pro binární klasifikaci $Y \in \{0, 1\}$, na kterou se omezíme, tedy vrací pravděpodobnost 1, $P(Y = 1) \in [0, 1]$.

0.1 Použití pro binární klasifikaci

0.1.1 Sigmoida

V modelu použijeme lineární výraz $\boldsymbol{w}^T\boldsymbol{x} = w_0 + w_1x_1 + \ldots + w_px_p$, který má však obor hodnot na celém $\mathbb{R}$. Tento výraz proto dosadíme do funkce, která je ostře rostoucí a má obor hodnot podmnožinu $[0, 1]$. V logistické regresi volíme sigmoidu

$$f(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}$$

Jedná se o speciální případ logistické funkce

$$f(x) = \frac{L}{1 + e^{-k(x - x_0)}},$$

se supremem $L = 1$, koeficientem růstu $k = 1$ a středem $x_0 = 0$.

Sigmoida má jako definiční obor celé $\mathbb{R}$ a obor hodnot $(0, 1)$. Na celém definičním oboru je ostře rostoucí, limita pro $x \to -\infty$ je 0 a pro $x \to +\infty$ je 1. Také platí, že střed má v bodě 0: $f(0) = \frac{1}{2}$.

0.1.2 Fungování modelu logistické regrese

Binární klasifikace vysvětlované proměnné $Y \in \{0, 1\}$ s p příznaky $X_1, \ldots, X_p$ logistická regrese provede predikcí pravděpodobnosti

$$\mathrm{P}(Y = 1 \mid \boldsymbol{x}, \boldsymbol{w}) = \frac{e^{\boldsymbol{w}^T\boldsymbol{x}}}{1 + e^{\boldsymbol{w}^T\boldsymbol{x}}},$$

kde $\boldsymbol{x} = (1, x_1, \ldots, x_p)$ je vektor hodnot příznaků a $\boldsymbol{w} = (w_0, \ldots, w_p)$ je vektor koeficientů. Model zvolí 1 když $\mathrm{P}(Y = 1 \mid \boldsymbol{x}, \boldsymbol{w}) > 0.5$, jinak predikuje 0.

0.2 Hranice rozhodnutí

Hranice rozhodnutí je dána rovnicí

$$\mathrm{P}(Y = 1 \mid \boldsymbol{x}, \boldsymbol{w}) = 0.5 \quad \Leftrightarrow \quad \boldsymbol{w}^T\boldsymbol{x} = 0 \quad \Leftrightarrow \quad w_0 + w_1x_1 + \ldots + w_px_p = 0$$

To odpovídá nadrovině v prostoru $\mathbb{R}^p$. Hranici tedy tvoří lineární varieta dimenze $p - 1$.

0.3 Logistická regrese jako MLE odhad

Vzhledem k tomu, že u logistické regrese predikujeme pravděpodobnost hodnot proměnné Y, nelze vyloženě měřit chybu takových odhadů a následně je minimalizovat jako u lineární regrese. Proto parametry $\boldsymbol{w}$ odhadujeme MLE (maximum likelihood estimation) metodou maximální věrohodností.

0.3.1 Myšlenka MLE odhadu

Pro parametry $\boldsymbol{w} = (w_0, w_1, \ldots, w_p)$ jsou pravděpodobnosti následující

$$
\begin{aligned}
p_1(\boldsymbol{x}; \boldsymbol{w}) = \mathrm{P}(Y = 1 \mid \boldsymbol{x}, \boldsymbol{w}) &= \frac{e^{\boldsymbol{w}^T \boldsymbol{x}}}{1 + e^{\boldsymbol{w}^T \boldsymbol{x}}} \\
p_0(\boldsymbol{x}; \boldsymbol{w}) = \mathrm{P}(Y = 0 \mid \boldsymbol{x}, \boldsymbol{w}) &= \frac{1}{1 + e^{\boldsymbol{w}^T \boldsymbol{x}}}
\end{aligned}
$$

MLE je takový odhad parametrů, pro které je daná realizace náhodného výběru nejpravděpodobnější (má největší věrohodnost). Metoda maximální věrohodnosti formálně odhaduje hodnotu $\hat{\boldsymbol{w}}$ parametru $\boldsymbol{w}$, která maximalizuje $L(\boldsymbol{w}; \boldsymbol{X})$ na trénovacích datech $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, kde $\boldsymbol{x}_i = (1, x_{i,1}, \ldots, x_{i,p})$ jsou jednotlivé naměřené hodnoty:

$$
\hat{\boldsymbol{w}} \in \left\{ \underset{\boldsymbol{w} \in \mathbb{R}^{p+1}}{\arg\max}\, L(\boldsymbol{w}; \boldsymbol{X}) \right\}, \quad \text{kde } L(\boldsymbol{w}; \boldsymbol{X}) = \prod_{i=1}^{N} p_{Y_i}(\boldsymbol{x}_i; \boldsymbol{w})
$$

0.3.2 Sestavení optimalizační úlohy pro trénování

Pro tuto funkci se snažíme nalézt maximum (nemusí existovat). Před zderivováním se však často vyplatí věrohodnost zlogaritmovat (log-likelihood), který je na $(0, +\infty)$ prostý a ostře rostoucí a tudíž má maximum ve stejném bodě.

$$
\begin{aligned}
\ell(\boldsymbol{w}; \boldsymbol{X}) &= \ln L(\boldsymbol{w}; \boldsymbol{X}) = \sum_{i=1}^{N} \ln p_{Y_i}(\boldsymbol{x}_i; \boldsymbol{w}) \\
&= \sum_{i=1}^{N} Y_i \ln p_1(\boldsymbol{x}_i; \boldsymbol{w}) + (1 - Y_i) \ln p_0(\boldsymbol{x}_i; \boldsymbol{w}) = \ldots \\
&= \sum_{i=1}^{N} \Big(Y_i \boldsymbol{w}^T \boldsymbol{x}_i - \ln\Big(1 + e^{\boldsymbol{w}^T \boldsymbol{x}_i}\Big)\Big)
\end{aligned}
$$

Parciální derivace a gradient vychází

$$
\frac{\partial \ell(\boldsymbol{w}; \boldsymbol{X})}{\partial w_j} = \sum_{i=1}^{N} \left(Y_i \boldsymbol{x}_{i,j} - \frac{e^{\boldsymbol{w}^T \boldsymbol{x}_i} \cdot \boldsymbol{x}_{i,j}}{1 + e^{\boldsymbol{w}^T \boldsymbol{x}_i}} \right) = \sum_{i=1}^{N} \boldsymbol{x}_{i,j} \left(Y_i - p_1(\boldsymbol{x}; \boldsymbol{w}) \right)
$$

$$
\nabla \ell(\boldsymbol{w}; \boldsymbol{X}) = \boldsymbol{X}^T \left(\boldsymbol{Y} - \boldsymbol{P} \right), \quad \text{kde } \boldsymbol{P} = (p_1(\boldsymbol{x}_1; \boldsymbol{w}), \ldots, p_1(\boldsymbol{x}_N; \boldsymbol{w}))^T
$$

Náš odhad leží v bodě, kde věrohodnost nabývá maxima, což nalezneme položením gradientu nule:

$$\nabla \ell(\hat{\boldsymbol{w}}; \boldsymbol{X}) = \boldsymbol{X}^T \left(\boldsymbol{Y} - \hat{\boldsymbol{P}} \right) = \mathbf{0}$$

Zde neexistuje explicitní řešení a je třeba jej hledat numerickými aproximativními metodami (např. vícerozměrnou Newtonovou metodou lze ukázat, že řešení konverguje k lok. maximu, které je v případě logistické regrese současně globálním maximem).

Výpočet koeficientů logistické regrese je výpočetně náročný. Bez explicitního vzorce je výsledek jen aproximace a je možné, že počítač nic nevrátí (nepodaří se mu nalézt dostatečně dobrá aproximace).

Funkce $\ell(\boldsymbol{w}; \boldsymbol{X})$ také žádné maximum mít nemusí. V takovém případě se numerickou metodou pouze snažíme hledat přibližné řešení $\boldsymbol{X}^T \left(\boldsymbol{Y} - \hat{\boldsymbol{P}} \right) = \mathbf{0}$.