16. LOGISTICKÁ REGRESE

Použití pro binární klasifikaci (jiné ani není!)

- rozhodnutí se konstruuje pomocí lineární kombinace příznaků (aka lin. regrese)

$$w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$$

- TRIK: místo hodnoty vysvětlované proměnné $Y \in \{0,1\}$
  - ↪ predikujeme pravděpodobnost, že $P(Y=1)$ z intervalu $[0,1]$
  - ↪ značíme $P(Y=1 \mid x, w)$ (vyjádřeno, že Y je závislá na hodnotách příznaků $x=(x_1,x_2,x_3)$ a $w=(w_0,w_1,w_2,w_3)$)
  - → platí $P(Y=1 \mid x,w) = 1 - P(Y=0 \mid x,w)$

- TRIK: potřebujeme, aby $w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$ bylo v intervalu $[0,1]$
  - ↪ dosadíme do vhodně zvolené funkce: sigmoida

- logistická funkce sigmoid: (obor hodnot: $(0,1)$)

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

$$\lim_{x\to\infty} f(x) = 1 \qquad \lim_{x\to-\infty} f(x) = 0$$

$f(x)$ … pravděpodobnost jevu $Y=1$

$1-f(x)$ … opačný jev

- přidáme X_0, který vždy $x_0 = 1$ (intercept) = očekávaná hodnota při nulových příznacích (platí i pro w_0)
- popis modelu: binární vysvětlovaná proměnná Y
  - p příznaků $X_1, X_2, \dots, X_p$
  - konstantní $X_0 = 1$
  - hledáme model ve tvaru

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

- predikce: pro daná x spočítáme odhad $P(Y=1 \mid x,w)$
  - ↪ pokud $> \frac{1}{2} \Rightarrow$ **Y=1**
  - jinak $\Rightarrow$ **Y=0**

- hranice rozhodnutí: hranici hledáme jako

$$P(Y=1 \mid x,w) = \frac{e^{w^T x}}{1+e^{w^T x}} = \frac{1}{2} \Rightarrow w^T x = 0 \leftarrow \text{nadrovina v } \mathbb{R}^p$$

(lin. varieta dim $p-1$)

UČENÍ MODELU

- používá se MLE (maximálně věrohodný odhad)
  ↳ maximalizujeme pravděpodobnost na základě trénovacích dat
- MLE (příklad na hodu mincí - tzn. 1 parametr p)
  Y.... náhodná veličina hodů (0,1)
  $p = P(Y=1)$ ⇒ chceme na základě měření (trénovacích dat) odhadnout p
  trénovací data: $7\times\ Y=1$, $3\times\ Y=0$ } celkem 10 hodů
  umíme spočítat s jakou pravděpodobností padnou ~~tes~~ trénovací data pro zvolené p:
  - např. pro $p=\frac{1}{2}$: $\binom{10}{7}\left(\frac{1}{2}\right)^{10} = 0{,}117$
    pro $p=0{,}6$: $\binom{10}{7}(0{,}6)^7(1-0{,}6)^3 = 0{,}215$ pravděpodobnější (lepší p)
    ⋮
  ⇒ vede na hledání maxima funkce
  $$L(p) = \binom{10}{7}p^7(1-p)^3 \quad \text{(na intervalu } \langle 0,1\rangle)$$
  1) zahodíme konstantu $\binom{10}{7}$ (logaritmus ostře rostoucí fce - nevadí)
  2) funkci zlogaritmujeme
  $$= \ln p^7(1-p)^3 = 7\ln p + 3\ln(1-p) = l(p)$$
  3) derivujeme $l(p)$:
  $$l'(p) = \frac{7}{p} - \frac{3}{1-p}$$
  4) spočítáme nulový bod:
  $$\hat{p}^{(MLE)} = \frac{7}{10}$$
- pro $p=\frac{7}{10}$: $\binom{10}{7}0{,}7^7\cdot(1-0{,}7)^3 = 0{,}27$ ← maximální na trénovacích datech
  ↑ MLE (volba $p=\frac{7}{10}$ má důvěryhodnost)
- MLE odhad pro logistickou regresi
  - podobné jako pro minci (2 hodnoty: $Y=1$; $Y=0$)
  × máme však více parametrů $w_0, w_1, \dots, w_p$
  $$p_1(x,w) = P(Y=1|x,w) = \boxed{\frac{e^{w^Tx}}{1+e^{w^Tx}}}$$
  $$p_0(x,w) = P(Y=0|x,w) = 1 - \frac{e^{w^Tx}}{1+e^{w^Tx}} = \boxed{\frac{1}{1+e^{w^Tx}}}$$
  - pravděpodobnost 1 datového bodu: $p_{Y_i}(x_i, w)$
  - máme N datových bodů x (s příznaky $x_0, x_1, \dots, x_p$)
  ↳
  $$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ \vdots & & & \ddots & \vdots \\ 1 & x_{N,1} & x_{N,2} & \cdots & x_{N,p} \end{pmatrix}$$
  ↑ intercept
  Y_i jsou vysvětlované proměnné
- předpokládáme, že data jsou nezávislá ⇒ celkovou pravděpodobnost zapíšeme jako součin
  $$\boxed{L(w) = \prod_{i=1}^{N} p_{Y_i}(x_i, w)}$$
  → tuto funkci se snažíme MAXIMALIZOVAT

– místo derivace součinu budeme derivovat logaritmus:

$$l(w) = \ln L(w) = \sum_{i=1}^{N} \ln p_{Y_i}(x_i, w) =$$

$$= \sum_{i=1}^{N} \left(Y_i \ln p_1(x_i, w) + (1 - Y_i) \ln p_0(x_i, w) \right)$$

$$= \sum_{i=1}^{N} \left(Y_i \ln\left(\frac{e^{w^T x}}{1 + e^{w^T x}} \right) + (1 - Y_i) \ln\left(\frac{1}{1 + e^{w^T x}} \right) \right) =$$

MAGIC (i v přednášce vynecháno)

$$= \sum_{i=1}^{N} \left(Y_i w^T x - \ln\left(1 + e^{w^T x}\right) \right)$$

$\Downarrow$

chceme najít maximum

$\Downarrow$

gradient (vektor složený z parciálních derivací podle všech proměnných $w_0, w_1, \dots, w_p$)

$$\frac{\partial l}{\partial w_j}(w) = \sum_{i=1}^{N} \left(x_{ij} \left(Y_i - p_1(x_i, w) \right) \right)$$

pro $j = 0, 1, \dots, p$

$\Downarrow$

$\nabla l(w) = X^T (Y - P)$ kde $P = (p_1(x_1, w), p_1(x_2, w), \dots, p_1(x_N, w))^T$

$\Downarrow$

$$\boxed{\nabla l(w) = X^T (Y - P) = 0}$$

konečný! (sigmoidy a exponenciály)

$\Rightarrow$ NEUMÍME NAJÍT EXPLICITNÍ ŘEŠENÍ

$\Downarrow$

numerické aproximativní metody (např. grad. sestup)