7.3 Generativni modeli

- u log. regresiji direktno modeliramo $p_k(x)$, $k \in S$
  $\longrightarrow$ primjer "diskriminacijske" metode
- alternativno, ako modeliramo $\boxed{f_k(x) := \text{"} \mathbb{P}(X = x \mid Y = k)}$, $x \in \mathbb{R}^p$, $k \in S$, te $\boxed{\pi_k := \mathbb{P}(Y = k)}$,

$$\boxed{\underset{\text{Bayes}}{p_k(x)} = \frac{f_k(x)\pi_k}{\sum_{i \in S} f_i(x)\pi_i}} \qquad (7.10)$$

$\longrightarrow$ zapravo modeliramo $\boxed{\mathbb{P}(X = x, Y = k) = f_k(x)\pi_k}$, tj. $\mathbb{P}((X, Y) \in \cdot) \longrightarrow$ "generativne" metode

$\longrightarrow$ 1° Linearna diskriminacijska analiza (LDA)
2° Kvadratna ——— " ——— (QDA)
3° Naivni Bayes

Uočimo, (7.10) povlači (uz 0-1 gubitak)

$$f^*(x) = \underset{k \in S}{\operatorname{argmax}} \boxed{f_k(x)\pi_k}. \qquad (7.11)$$

7.3.1 LDA

$\longrightarrow$ Pretpostavka je da $X \mid Y = k \sim N(\underset{\in \mathbb{R}^p}{\mu_k}, \underset{\in \mathbb{R}^{p \times p}}{\Sigma_k})$, tj.

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}}\sqrt{\det(\Sigma_k)}} \cdot \exp\left\{-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right\}, \ x \in \mathbb{R}^p, \qquad (7.12)$$

te dodatno

$$\boxed{\Sigma_k = \Sigma, \ \forall k \in S \, !} \qquad (7.13)$$

(Mahalanobisova udaljenost od x i μ_k)

(7.14)

$$\hookrightarrow t^*(x) = \arg\max_{k\in S} \log(f_k(x)\cdot \pi_k)$$

$$= \log(f_k(x)) + \log(\pi_k)$$

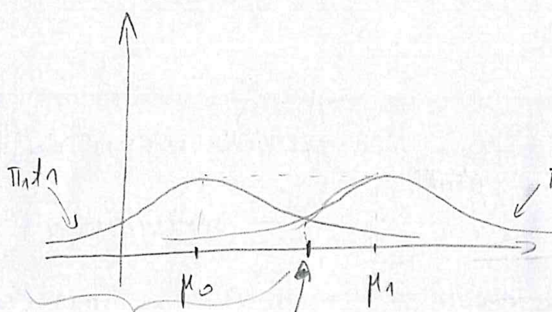
$$\overset{(7.12)}{\underset{(7.13)}{=}} \arg\min_{k\in S} \Big\{ \underbrace{\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k) - \log(\pi_k)}_{=: d_k(x)} \Big\}, \; x \in \mathbb{R}^p$$

Ako je $\boxed{\Sigma = I_p}$ te $\pi_0 = \dots = \pi_{K-1}$,

$$\boxed{t^*(x) = \arg\min_{k\in S} \|x-\mu_k\|_2^2 .}$$

Pr. $p=1$, $K=2$ $\quad (\Sigma = \sigma^2 > 0)$

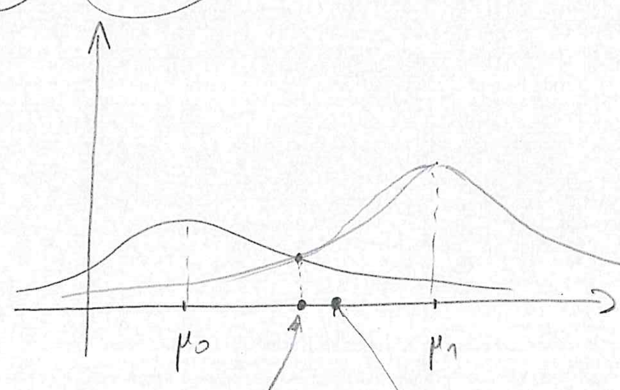
1° $\pi_0 = \pi_1$



$t^*(x) = 0$

$\frac{\mu_0+\mu_1}{2}$ = granica odluke

2° $\pi_1 > \pi_2$



granica odluke $< \frac{\mu_0+\mu_1}{2}$

[uzimamo u obzir da je više (x_i, y_i) s $y_i = 1$] ■

Iz $d_k(x) = \frac{1}{2}x^T\Sigma^{-1}x - x^T \underbrace{\Sigma^{-1}\mu_k}_{=: w_k} + \underbrace{\mu_k^T \Sigma^{-1}\mu_k - \log(\pi_k)}_{=: -b_k}$

← ne ovisi o k

$$\hookrightarrow t^*(x) = \arg\max_{k\in S} \; x^T w_k + b_k$$

↳ linearne diskriminacijske f-je ⟶ "LDA"

Specijalno, LDA ⟶ <u>linearne</u> granice odluke.

Parametre procjenjujemo s:

- $\hat{\pi}_h = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{y_i = h\}}$,
- $\hat{\mu}_h = \dfrac{\sum_{i=1}^{n} x^{(i)} \mathbb{1}_{\{y_i=h\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{y_i=h\}}}$,

"pooled" uzoračka kov. matrica

- $\hat{\Sigma} = \frac{1}{n-K} \sum_{i=1}^{n} (x^{(i)} - \hat{\mu}_{y_i})(x^{(i)} - \hat{\mu}_{y_i})^T$.

$\hookrightarrow \hat{f}(x) = \underset{h \in S}{\operatorname{argmin}}\ \hat{d}_h(x)$.

$(K-1)$ - dim. afin prostor

Nap. 1 (!) (i) Pretp. da je $\boxed{\hat{\Sigma} = I_p}$, te $H_{K-1} = \mu_0 \oplus \operatorname{span}\{\mu_1, \dots, \mu_{K-1}\}$.

$$\hookrightarrow \hat{f}(x) = \underset{h \in S}{\operatorname{argmin}} \left\{ \frac{1}{2} \| x - \mu_h \|_2^2 - \log(\pi_h) \right\}$$

$$= \underset{h \in S}{\operatorname{argmin}} \left\{ \frac{1}{2} \| \operatorname{Proj}_{H_{K-1}}(x) - \mu_h \|_2^2 - \log(\pi_h) \right\}$$

$\Rightarrow$ čim je $K-1 < p$, imamo smanjenje dimenzije

(ii) općenito, prvo napravimo transformaciju

$$x \longmapsto x^* := \hat{\Sigma}^{-\frac{1}{2}} x \quad \left(\Rightarrow (x-\hat{\mu}_h)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_h) = \| x^* - \hat{\mu}_h^* \|_2^2,\ \hat{\Sigma}^* = I_p \right)$$

te gledamo projekcije na $H_{K-1}^* = \mu_0^* \oplus \operatorname{span}\{\mu_1^*, \dots, \mu_{K-1}^*\}$.

$\hookrightarrow$ Može se pokazati da postoje vektori $\boxed{v_1, \dots, v_{K-1} \in \mathbb{R}^p}$

t. d.

$$\hat{f}(x) = \underset{h \in S}{\operatorname{argmin}} \left\{ \frac{1}{2} \| V \cdot x - V \cdot \hat{\mu}_h \|_2^2 - \log(\pi_h) \right\}$$

za $V = \begin{bmatrix} - v_1^T - \\ \vdots \\ - v_{K-1}^T - \end{bmatrix} \in \mathbb{R}^{(K-1) \times p}$

$\hookrightarrow$ $v_i^T \cdot x$ je tzv. i-ta diskriminacijska koordinata

$\hookrightarrow$ Primijenjena statistika i ESL 4.3.3 $\longrightarrow$ $\boxed{\text{wine-LDA.R}}$

(7.3.2) QDA

↳ sve kao kod LDA, ali **bez** pretp. $\Sigma_k = \Sigma,\ \forall k \in S$.

$$\Rightarrow f^*(x) \overset{pt}{=} \underset{k \in S}{\operatorname{argmax}} \left\{ -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k) - \frac{1}{2}\log(\det(\Sigma_k)) + \log \pi_k \right\}$$

$$= \underset{k \in S}{\operatorname{argmax}} \left\{ -\frac{1}{2} x^T \Sigma_k^{-1} x + \dots \right\}$$

↳ **kvadratne** granice odluke.

Ipak, u odnosu na LDA, QDA procjenjuje $\hat{\Sigma}_k, k \in S$

$\rightsquigarrow$ dodatnih $\frac{p(p-1)}{2} \cdot (K-1)$ parametara!

Kada je p **velik**, čak i LDA procjenjuje previše parametara

$$\rightsquigarrow \underbrace{\Sigma}_{\frac{p(p-1)}{2}},\ \underbrace{\mu_k}_{K \cdot p},\ \underbrace{\pi_k}_{K-1}$$

(7.3.3) Naivni Bayes (NB)

↳ pretp. da su $\forall k \in S$, $X_1, \dots, X_p$ **nezavisne** uz dano $Y = k$, tj.

$$\boxed{f_k(x) = \prod_{j=1}^{p} f_{k,j}(x_j)}, \quad \forall x = (x_1, \dots, x_p) \in \mathbb{R}^p \qquad (7.15)$$

[nezavisnost je oblik regularizacije]

Marginalne gustoće možemo procijeniti npr.

- ako $X_j \in \mathbb{R}$ ⇒ pretpostavimo

$$X_j \mid Y = k \sim N(\mu_{kj}, \sigma^2_{kj})$$

(kao QDA uz Σ_k = dijagonalna matrica)

- ako $x_j \in S_j = \{0, \dots, K_j - 1\}$ $\Rightarrow$ npr. neparametarski

$$\hat{f}_{k,j}(x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{y_i = k,\, x_{ij} = x\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{y_i = k\}}}.$$

7.3.4 Usporedba metoda

- log. regresija — $\log\left(\frac{p_k(x)}{p_0(x)}\right) = \boxed{x^T \beta^{(k)}}$ (tipično $x_0 = 1$)

- LDA — $\log\left(\frac{p_k(x)}{p_0(x)}\right) \overset{(Dz)}{=} \log \frac{\pi_k}{\pi_0} - \frac{1}{2}(\mu_k + \mu_0)^T \Sigma^{-1} (\mu_k - \mu_0)$

$$+ x^T \Sigma^{-1} (\mu_k - \mu_0)$$

$$=: \alpha_{k,0} + \sum_{i=1}^{K-1} \alpha_{k,i} \cdot x_i$$

$$=: \boxed{x^T \alpha^{(k)}}$$

$x_0 := 1$

$\hookrightarrow$ ipak, $\hat{\alpha}^{(k)} \neq \hat{\beta}^{(k)}$ [, te je log. regresija općenitija jer ne pretpostavlja mišće na razd. $\mathbb{P}(X \in \cdot)$].

- NB — $\log\left(\frac{p_k(x)}{p_0(x)}\right) \overset{(7.15)}{=} \log\left(\frac{\pi_k}{\pi_0}\right) + \sum_{j=1}^{p} \log\left(\frac{f_{kj}(x_j)}{f_{0j}(x_j)}\right)$, gdje je $p_k(x) \propto \pi_k f_k(x)$

$$=: \alpha_k + \sum_{j=1}^{p} g_{kj}(x_j)$$

$\hookrightarrow$ fleksibilne granice odluke

$\hookrightarrow$ ipak, nema interakcija

- QDA — $\log\left(\frac{p_k(x)}{p_0(x)}\right) = a_k' + \sum_{j=1}^{p} b_{kj} x_j + \boxed{\sum_{j,l=1}^{p} c_{kjl} x_j x_l}$

interakcije.

7.4 kNN metoda

↳ za $k_0 \in \{1, \dots, n\}$,

$$\hat{p}_k(x) = \frac{1}{k_0} \sum_{x^{(i)} \in N_{k_0}(x)} \mathbb{1}\{y_i = k\}, \quad k \in S,\ x \in \mathbb{R}^p$$

↳ $\hat{f}(x) = \underset{k \in S}{\arg\max}\, \hat{p}_k(x) = \underset{k \in S}{\arg\max} \sum_{x^{(i)} \in N_{k_0}(x)} \mathbb{1}\{y_i = k\}$

"majority vote"
[uz 0-1 gubitak]

Ncp. | za velike p
→ problem dimenzionalnosti!

→ $\boxed{S\text{-and-}P \cdot R}$