DORA OMANOVIĆ
0036522607

VIU 1.4

(a) $\hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} \vec{x}^{i} = (5.483, -0.183, -0.733)$

$$\hat{\Sigma}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} (\vec{x}^{i} - \hat{\mu}_{MLE})(\vec{x}^{i} - \hat{\mu}_{MLE})^T$$

$$= \begin{bmatrix} 8.77 & -1.198 & -4.792 \\ -1.198 & 0.191 & 0.766 \\ -4.792 & 0.766 & 3.062 \end{bmatrix}$$

(u Pythonu)

(b) $\vec{x} = (-2, 1, 0)$

$$p(\vec{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})\right)$$

$\boxed{|\Sigma| = 0}$!

→ Σ nema inverz i ne možemo dijeliti s 0

→ gustoća $p(\vec{x}|\mu,\Sigma)$ nije dobro definirana

(c) $$\rho_{x_1x_2} = \frac{\sigma_{x_1x_2}}{\sqrt{\sigma_{x_1}^2}\sqrt{\sigma_{x_2}^2}} = \frac{-1.198}{\sqrt{8.77 \cdot 0.191}} = -0.926$$

$$\rho_{x_1x_3} = \frac{\sigma_{x_1x_3}}{\sqrt{\sigma_{x_1}^2}\sqrt{\sigma_{x_3}^2}} = \frac{-4.792}{\sqrt{8.77 \cdot 3.062}} = -0.925$$

$$\rho_{x_2x_3} = \frac{\sigma_{x_2x_3}}{\sqrt{\sigma_{x_2}^2}\sqrt{\sigma_{x_3}^2}} = \frac{0.766}{\sqrt{0.191 \cdot 3.062}} = 1.002$$

→ izbaciti x_3 (sve su korelirane, može bilo koja)

$\mu' = (5.483, -0.183)$

$$\Sigma' = \begin{bmatrix} 8.77 & -1.198 \\ -1.198 & 0.191 \end{bmatrix}$$

$|\Sigma'| = 0.243$

$$(\Sigma')^{-1} = \begin{bmatrix} 0.786 & 4.922 \\ 4.922 & 36.04 \end{bmatrix}$$

$$p(\vec{x}'|\mu',\Sigma') = \frac{1}{(2\pi)^{\frac{2}{2}}|\Sigma|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}\underbrace{(\vec{x}'-\vec{\mu}')(\Sigma')^{-1}(\vec{x}'-\vec{\mu}')^T}_{7.32}\Big)$$

$\vec{x}' = (-2, 1)$

$n' = 2$

$p(\vec{x}'|\vec{\mu}',\Sigma') = 0.012$

VAU 1.5

DORA OMANOVIĆ
0036522607

(a) MAP procjenitelj:

$$\hat{\vec{\theta}}_{MAP} = \underset{\vec{\theta}}{\mathrm{argmax}}\; p(\vec{\theta}|D) = \underset{\vec{\theta}}{\mathrm{argmax}}\; p(D|\vec{\theta})\,p(\vec{\theta})$$

→ bolji od MLE-a jer kombinira vjerojatnost oznaka odnosno izglednost parametara $\vec{\theta}$ s apriornom distribucijom parametara, što nam omogućava da nađemo balans između apriornog znanja o parametrima i znanja dobivenog iz podataka

(b) Konjugatne distribucije su distribucije iste vrste, a konjugatna apriorna distribucija za funkciju izglednosti znači da ona pomnožena s funkcijom izglednosti daje konjugatnu distribuciju.

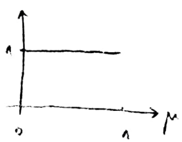
To nam je bitno da možemo raditi "online" učenje.

(c) $\mu \sim B(\alpha, \beta)$

$$p(\mu|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\,\mu^{\alpha-1}(1-\mu)^{\beta-1}$$

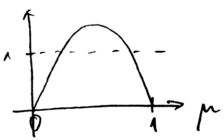
(1) $\alpha = \beta = 1$

$p(\mu) = 1$



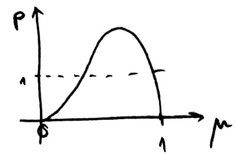
(2) $\alpha = \beta = 2$

$p(\mu|2,2) = \frac{1}{B(2,2)}\mu(1-\mu)$



(3) $\alpha = 2, \beta = 4$

(4) $\alpha = 4, \beta = 2$



(d) $p(\mu|N,m,\alpha,\beta) = ?$

Bayes: $p(\mu|D) = \dfrac{p(D|\mu)\,p(\mu)}{p(D)}$

izglednost μ

$p(D|\mu) = \mu^m(1-\mu)^{N-m}$

apriorna

$p(\mu|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\mu^{\alpha-1}(1-\mu)^{\beta-1}$

$$p(\mu|N,m,\alpha,\beta) = \mu^m(1-\mu)^{N-m}\,\mu^{\alpha-1}(1-\mu)^{\beta-1}\,\frac{1}{B(\alpha,\beta)\,p(D)}$$

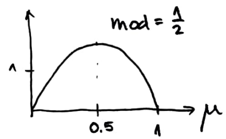
$$= \mu^{(m+\alpha)-1}(1-\mu)^{(N-m+\beta)-1}\cdot\frac{1}{B(\alpha,\beta)\,p(D)}$$

$$= \frac{1}{B(\alpha',\beta')}\,\mu^{\alpha'-1}(1-\mu)^{\beta'-1}$$

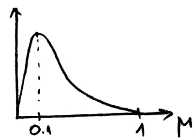
(e) $\alpha = \beta = 2$

$N = 10$

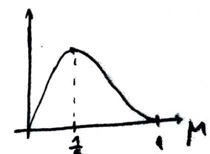
$m = 1$



apriorna
$p(\mu|\alpha,\beta)$



izglednost
$\mathcal{L}(\mu|N,m)$



umnožak

$\text{mod} = \frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{3-1}{3+11-2} = \frac{2}{12}$

(f) $\hat{\mu}_{MAP} = \frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{m+\alpha-1}{\cancel{m}+\alpha+N-\cancel{m}+\beta-2} = \frac{m+\alpha-1}{N+\alpha+\beta-2} = \frac{1}{6}$

↓ iz D i iz apriorne

$\hat{\mu}_{MLE} = \frac{m}{N} = \frac{1}{10}$ → samo iz D

Ako povećamo N, više uzimamo podatke u obzir i smanjuje se razlika između MAP i MLE.

(g) Ako $\alpha=\beta=2$, za $\hat{\mu}_{MAP}$ dobivamo

$\hat{\mu}_{MAP} = \frac{m+1}{N+2}$ Laplaceov procjenitelj.

$\boxed{\text{DORA OMANOVIĆ} \\ 0036522607}$

$\boxed{\text{V15 1.3}}$

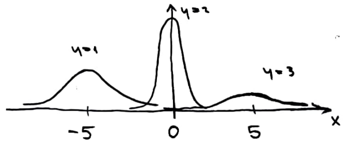
$X = \mathbb{R}$ $\quad P(y=1)=0.3 \quad \mu_1=-5 \quad \sigma_1^2=5$

$K=3$ $\quad P(y=2)=0.2 \quad \mu_2=0 \quad \sigma_2^2=1$

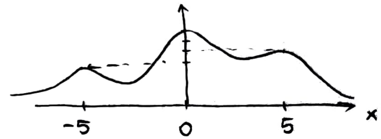
$\quad P(y=3)=0.5 \quad \mu_3=5 \quad \sigma_3^2=10$

$$P(y=j\mid\vec{x}) = \frac{p(\vec{x}\mid y=j)P(y=j)}{\sum_{y'} p(\vec{x}\mid y=y')P(y=y')}$$

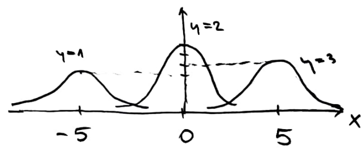
$p(x\mid y) \rightarrow$ Gauss



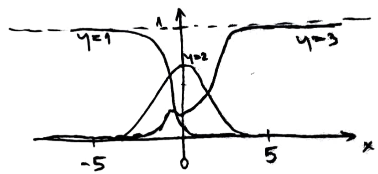
$p(x) = \sum p(x,y)$



$p(x,y) \rightarrow$ brojnik $\forall j$



$p(y\mid x) = \frac{p(x,y)}{p(x)}$



V15 1.5

DORA OMANOVIĆ
0036522607

$y \in \{1,2\}$

$\vec{x} = (\underbrace{x_1, x_2, x_3, x_4}_{\text{ocjene}}, \underbrace{x_5, x_6}_{\text{mature}}) \Rightarrow n=6$

H_1 dijeljena Σ

H_2 dijagonalna i dijeljena Σ

H_3 izotropna Σ

(a) parametri

$\vec{\mu}_1, \vec{\mu}_2, \Sigma, P(y=j)$

H_1: $\underbrace{n+n}_{n \cdot K} + \frac{n(n+1)}{2} + 1 = 34$

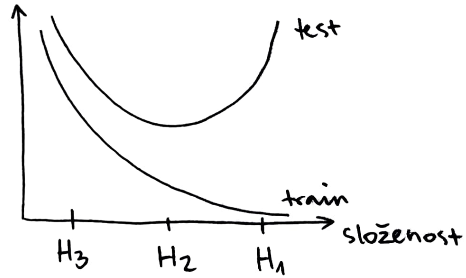
H_2: $n+n+n+1 = 3n+1 = 19$

H_3 $n+n+K \cdot 1 + 1 = 15$

(b) Koji najbolje generalizira?

Pretpostavljamo da su značajke vjerojatno međusobno zavisne, pogotovo $x_1 - x_4$ i $x_5 - x_6$.
H_1 modelira zavisnost značajki, međutim u podacima nije dobro imati nebitne značajke i H_1 bi mogao dovesti do prenaučenosti. H_3 nije dobar jer ima premalo parametara i bit će podnaučen, a H_2 bi onda najbolje generalizirao.

(c)



(d) Za odabir modela trebamo napraviti unakrsnu provjeru.
Treba izbaciti nebitne značajke (neke jako korelirane).

...distribuciju parametra, $p(\mu|N, m, \alpha, \beta)$

VJ6 1.2

DORA OMANOVIĆ
0036522607

(a) MLE procjene
Naivan Bayes

x_1 - mjesto
x_2 - otok
x_3 - smještaj
x_4 - prijevoz

$P(y=da) = \frac{4}{7}$

$P(y=ne) = \frac{3}{7}$

MJESTO:

$P(Istra|y=da) = \frac{2}{4} = \frac{1}{2}$ $\quad P(Istra|y=ne) = \frac{0}{3} = 0$ (*)

$P(Kvarner|y=da) = \frac{0}{4} = 0$ $\quad P(Kvarner|y=ne) = \frac{2}{3}$

$P(Dalm.|y=da) = \frac{2}{4} = \frac{1}{2}$ $\quad P(Dalm.|y=ne) = \frac{1}{3}$

OTOK:

$P(da|y=da) = \frac{3}{4}$ $\quad P(da|y=ne) = \frac{0}{3} = 0$ (□)

$P(ne|y=da) = \frac{1}{4}$ $\quad P(ne|y=ne) = \frac{3}{3} = 1$

SMJEŠTAJ

$P(priv|y=da) = \frac{2}{4} = \frac{1}{2}$ $\quad P(priv|y=ne) = \frac{1}{3}$

$P(kamp|y=da) = \frac{0}{4} = 0$ (*) $\quad P(kamp|y=ne) = \frac{2}{3}$

$P(hotel|y=da) = \frac{2}{4} = \frac{1}{2}$ $\quad P(hotel|y=ne) = \frac{0}{3} = 0$ (□)

PRIJEVOZ

$P(auto|y=da) = \frac{3}{4}$ $\quad P(auto|y=ne) = 0$

$P(bus|y=da) = 0$ (*) (□) $\quad P(bus|y=ne) = \frac{2}{3}$

$P(avion|y=da) = \frac{1}{4}$ $\quad P(avion|y=ne) = \frac{1}{3}$

$h(Istra, ne, kamp, bus) = \underset{y}{\operatorname{argmax}} P(y) \prod_j P(x_j|y)$

=> zbog (*) ispadne 0 $\forall y$

$h(Dalm., da, hotel, bus) = \underset{y}{\operatorname{argmax}} \ldots$

=> zbog (□) ispadne 0 $\forall y$

=> ne možemo klasificirati!

(b) Laplaceove procjene: $\frac{m+1}{N+x}$

broj razl. mogućih realizacija: za MJESTO, SMJEŠTAJ I PRIJEVOZ je 3, a za OTOK 2.

MJESTO \ y=	da	ne
Istra	$\frac{3}{7}$	$\frac{1}{6}$
Kvarner	$\frac{1}{7}$	$\frac{3}{6}$
Dalm.	$\frac{3}{7}$	$\frac{2}{6}$

OTOK \ y=	da	ne
da	$\frac{4}{6}$	$\frac{1}{5}$
ne	$\frac{2}{6}$	$\frac{4}{5}$

SMJEŠTAJ \ y=	da	ne
privatni	$\frac{3}{7}$	$\frac{2}{6}$
kamp	$\frac{1}{7}$	$\frac{3}{6}$
hotel	$\frac{3}{7}$	$\frac{1}{6}$

PRIJEVOZ \ y=	da	ne
auto	$\frac{4}{7}$	$\frac{1}{6}$
bus	$\frac{1}{7}$	$\frac{3}{6}$
avion	$\frac{2}{7}$	$\frac{2}{6}$

$h(Istra, ne, kamp, bus) = ne$

$\frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{7} \cdot \frac{1}{7} \cdot \frac{4}{7} = 1.666 \cdot 10^{-3}$ za $y=da$

$\frac{1}{6} \cdot \frac{4}{5} \cdot \frac{3}{6} \cdot \frac{3}{6} \cdot \frac{3}{7} = 0.014$ za $y=ne$

analogno:

$h(Dalmacija, da, hotel, bus) = da$ jer 0.01 za $y=da$ i 0.002 za $y=ne$