6. Bayesov klasifikator (nastavak)

prof. dr. sc. Bojana Dalbelo Bašić
doc. dr. sc. Jan Šnajder

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Ak. god. 2012/13.

v1.0

Danas. . .

1. Jednodimenzijski Bayesov klasifikator za kontinuirane ulaze

2. Višedimenzijski Bayesov klasifikator za kontinuirane ulaze

3. Napomene

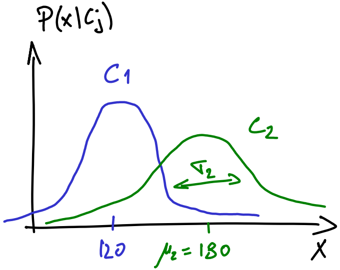
Danas. . .

1. Jednodimenzijski Bayesov klasifikator za kontinuirane ulaze
2. Višedimenzijski Bayesov klasifikator za kontinuirane ulaze
3. Napomene

Izglednosti klasa

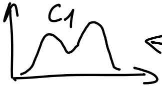
$$x|\mathcal{C}_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$$

$$p(x|\mathcal{C}_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ -\frac{(x-\mu_j)^2}{2\sigma_j^2} \right\}$$



Napomena!

Razdioba primjera unutar jedne klase je UNIMODALNA (modelirana jednim Gaussom)



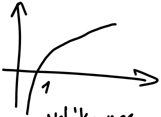
nije unimodalno! (mješavina Gauss. razdioba)

Gaussova

Multinomijalna

$$\boxed{h_j(x) = p(x|\mathcal{C}_j)P(\mathcal{C}_j)} = p(x,\mathcal{C}_j)$$

log



velike neg. vrijednosti

$$\Rightarrow \quad \ln p(x|\mathcal{C}_j) + \ln P(\mathcal{C}_j)$$

$$= -\frac{1}{2}\cancel{\ln 2\pi} - \ln \sigma_j - \frac{(x-\mu_j)^2}{2\sigma_j^2} + \ln P(\mathcal{C}_j)$$

konst.

pouzdanost klasifikacije

$$h''_j(x|\boldsymbol{\theta}_j) = -\ln \hat{\sigma}_j - \frac{(x-\hat{\mu}_j)^2}{2\hat{\sigma}_j^2} + \ln \hat{P}(\mathcal{C}_j)$$

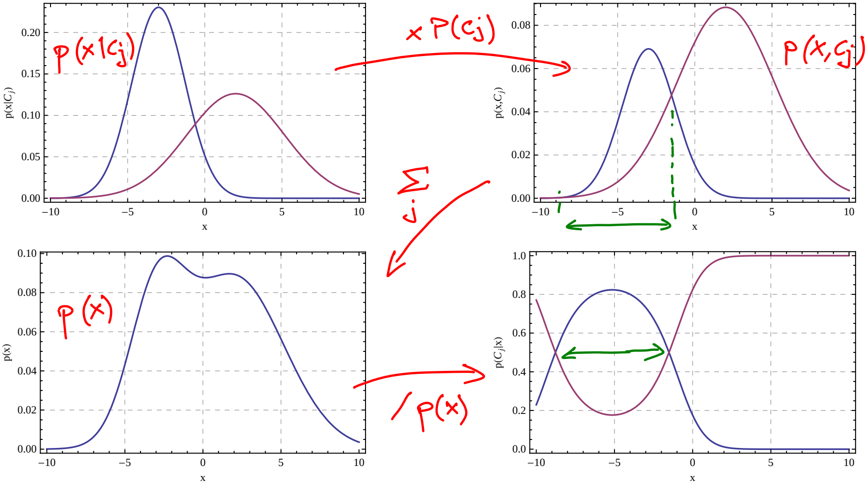
$$\boxed{\boldsymbol{\theta}_j = (\mu_j, \sigma_j, P(\mathcal{C}_j))}$$

MLE:

$$\hat{\mu}_j = \frac{1}{N_j}\sum_{i=1}^{N} \mathbf{1}\{y^{(i)} = \mathcal{C}_j\}x^{(i)} \qquad \hat{\sigma}_j^2 = \frac{1}{N_j}\sum_{i=1}^{N} \mathbf{1}\{y^{(i)} = \mathcal{C}_j\}(x^{(i)} - \hat{\mu}_j)^2 \qquad \hat{P}(\mathcal{C}_j) = \frac{N_j}{N}$$

Gustoće vjerojatnosti

$$p(x|\mathcal{C}_1) \sim \mathcal{N}(-3,3),\ p(x|\mathcal{C}_2) \sim \mathcal{N}(2,10),\ P(\mathcal{C}_1) = 0.3 \text{ i } P(\mathcal{C}_2) = 0.7$$



Danas...

1. Jednodimenzijski Bayesov klasifikator za kontinuirane ulaze
2. Višedimenzijski Bayesov klasifikator za kontinuirane ulaze
3. Napomene

Višedimenzijski Bayesov klasifikator

Izglednost klase:

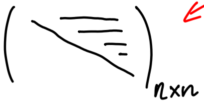
$$p(\mathbf{x}|\mathcal{C}_j) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_j|^{1/2}} \exp\Big\{ -\frac{1}{2}\overbrace{(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)}^{\Delta^2}\Big\}$$

Model:

$$\begin{aligned} h_j(\mathbf{x}) &= \ln p(\mathbf{x}|\mathcal{C}_j) + \ln P(\mathcal{C}_j) \\ &= \cancel{-\frac{n}{2}\ln 2\pi} - \frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) + \ln P(\mathcal{C}_j)\end{aligned}$$

konst.

Broj parametara?



$K \cdot n$

$K-1$

$K(n^2/2 + n/2) = \frac{n}{2}(n+1)K \Rightarrow O(n^2)$

Kovarijacijska matrica – napomene

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) & \dots & \mathrm{Cov}(x_1, x_n) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) & \dots & \mathrm{Cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(x_n, x_1) & \mathrm{Cov}(x_n, x_2) & \dots & \mathrm{Var}(x_n) \end{pmatrix}$$

Varijanca

- $\boldsymbol{\Sigma}$ je simetrična i ima sve elemente nenegativne
- Zato je $\boldsymbol{\Sigma}$ uvijek pozitivno semidefinitna: $\Delta^2 = \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{x} \geqslant 0$
  $\Rightarrow$ za Mahalanobisovu udaljenost vrijedi $\Delta \geqslant 0$
- Ali, da bi PDF bila dobro definirana, $\boldsymbol{\Sigma}$ mora biti pozitivno definitna: $\Delta^2 = \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{x} > 0$ za ne-nul vektor $\mathbf{x}$
- $\boldsymbol{\Sigma}$ je pozitivno definitna $\Rightarrow$ $\boldsymbol{\Sigma}$ je nesingularna: $|\boldsymbol{\Sigma}| > 0$ i postoji $\boldsymbol{\Sigma}^{-1}$ (obrat ne vrijedi!)
- Ako $\boldsymbol{\Sigma}$ nije pozitivno definitna, najčešći uzroci su $\mathrm{Var}(x_i) = 0$ (beskorisna značajka) ili $\mathrm{Cov}(x_i, x_j) = 1$ (redundantne značajke)

Nelinearnost modela

konst.

$$h_j(\mathbf{x}) = -\frac{n}{2}\cancel{\ln 2\pi} - \frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) + \ln P(\mathcal{C}_j)$$

$$\Rightarrow\ -\frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j) + \ln P(\mathcal{C}_j)$$

kvadratni član

$$\longrightarrow (x_1\ \ x_2)\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$a x_1^2 + (b+c)x_1 x_2 + d x_2^2$$

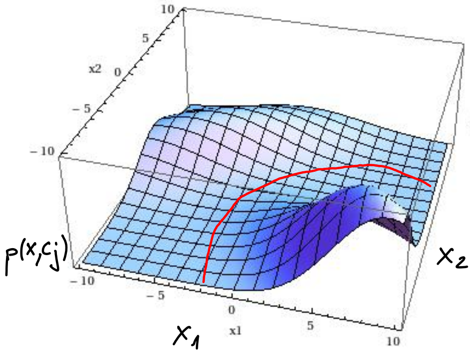
Granica između klasa C_1 i C_2:

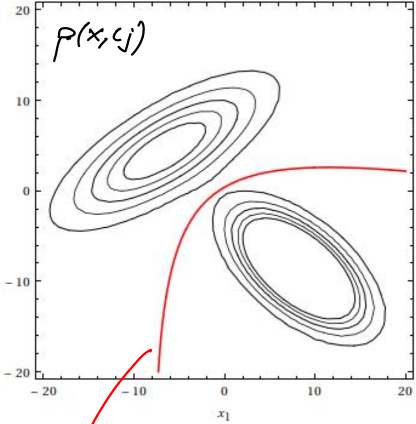
$h_1(x) = h_2(x)$

$\Rightarrow h_1(x) - h_2(x) = 0$

Nelinearnost modela

$n=2$





$h_1(x) = h_2(x)$

Procjena parametara

MLE:

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_j &= \frac{1}{N_j} \sum_{i=1}^{N} \mathbf{1}\{y^{(i)} = \mathcal{C}_j\}\mathbf{x}^{(i)} \\
\hat{\boldsymbol{\Sigma}}_j &= \frac{1}{N_j} \sum_{i=1}^{N} \mathbf{1}\{y^{(i)} = \mathcal{C}_j\}(\mathbf{x}^{(i)} - \hat{\boldsymbol{\mu}}_j)(\mathbf{x}^{(i)} - \hat{\boldsymbol{\mu}}_j)^{\mathrm{T}} \\
\hat{P}(\mathcal{C}_j) &= \frac{N_j}{N}
\end{aligned}
$$

Kvadratni model ima previše parametara: $\mathcal{O}(n^2)$

Pojednostavljenja $\Rightarrow$ dodatne induktivne pretpostavke?

1. pojednostavljenje: dijeljena kovarijacijska matrica

(shared cov. matrix)

$$\hat{\mathbf{\Sigma}} = \sum_j \hat{P}(\mathcal{C}_j)\hat{\mathbf{\Sigma}}_j$$

konst. konst. za zadani x

$$h_j(\mathbf{x}) = \cancel{-\frac{1}{2}\ln|\mathbf{\Sigma}|} - \frac{1}{2}(\cancel{\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{x}} - 2\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_j) + \ln P(\mathcal{C}_j)$$

$$\Rightarrow \quad \mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_j - \frac{1}{2}\boldsymbol{\mu}_j^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_j + \ln P(\mathcal{C}_j)$$

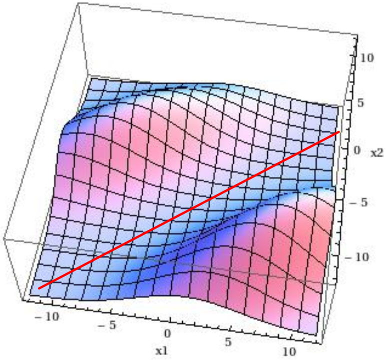
Linearan model! !

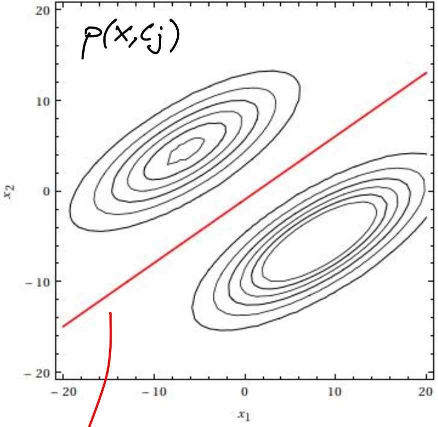
Broj parametara?

$\frac{n}{2}(n+1)$ $\quad nK \quad K-1$

$\Rightarrow O(n^2) \quad$:(

1. pojednostavljenje: dijeljena kovarijacijska matrica





$h_1(x) = h_2(x)$

2. pojednostavljenje: dijagonalna kovarijacijska matrica

$$\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i^2) \quad \Rightarrow \quad |\boldsymbol{\Sigma}| = \prod_i \sigma_i, \quad \boldsymbol{\Sigma}^{-1} = \mathrm{diag}(1/\sigma_i^2)$$

$$\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$$

Izglednost klase:

$$\begin{aligned} p(\mathbf{x}|\mathcal{C}_j) &= \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right\} \\ &= \frac{1}{(2\pi)^{n/2}\prod_{i=1}^{n}\sigma_i} \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_{ij}}{\sigma_i}\right)^2\right\} \\ &= \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{ -\frac{1}{2}\left(\frac{x_i-\mu_{ij}}{\sigma_i}\right)^2\right\}}_{\text{univarijatni gauss}} = \prod_{i=1}^{n} \mathcal{N}(\mu_{ij}, \sigma_i^2) \end{aligned}$$

$\Rightarrow$ Naivan Bayesov klasifikator: $p(x|C_j) = \prod_{i=1}^{n} p(x_i|C_j)$

2. pojednostavljenje: dijagonalna kovarijacijska matrica

Model:

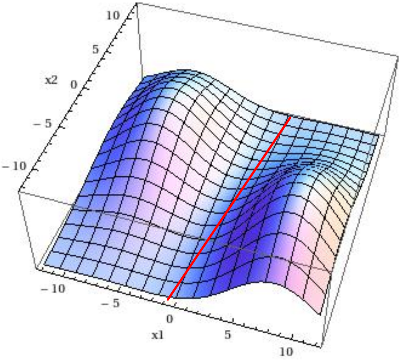
$$h_j(\mathbf{x}) = \ln p(\mathbf{x}|\mathcal{C}_j) + \ln P(\mathcal{C}_j)$$
$$\Rightarrow \; -\frac{1}{2}\underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\mu_{ij}}{\sigma_i}\right)^2}_{\text{normirana eukl. udaljenost}^2} + \ln P(\mathcal{C}_j)$$

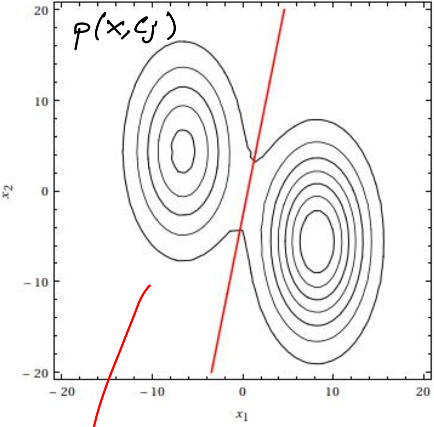
Naivan Bayesov klasifikator $\Rightarrow \; x_i \perp x_k | \mathcal{C}_j \;\Rightarrow\; \mathrm{Cov}(x_i|\mathcal{C}_j, x_k|\mathcal{C}_j) = 0$

Broj parametara?

$n + nK + K - 1 \quad \Rightarrow \; O(n)$:)

2. pojednostavljenje: dijagonalna kovarijacijska matrica





$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$$

elipse s osima paralelnima s x i y

3. pojednostavljenje: izotropna kovarijacijska matrica

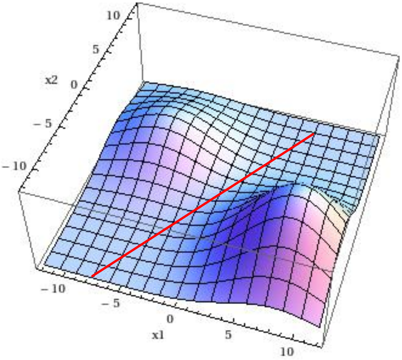
$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I} = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$$

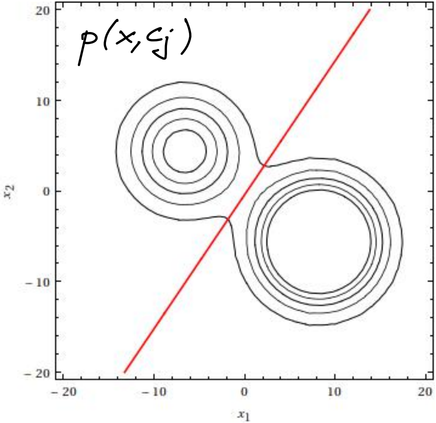
$$h_j(\mathbf{x}) = -\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{n} (x_i - \mu_{ij})^2}_{\text{euklidska udaljenost}^2} + \ln P(\mathcal{C}_j)$$

Broj parametara?

$(1+)\, Kn + K - 1 \Rightarrow O(n)$

3. pojednostavljenje: izotropna kovarijacijska matrica





4. pojednostavljenje: jednake apriorne vjerojatnosti

konst. konst.

$$h_j(\mathbf{x}) = -\frac{1}{\cancel{2\sigma^2}}\sum_{i=1}^{n}(x_i - \mu_{ij})^2 + \cancel{\ln P(\mathcal{C}_j)}$$

$\|x\| = \sqrt{x^T x}$

$$\Rightarrow \; -\|\mathbf{x} - \boldsymbol{\mu}_j\|^2$$

konst.

$$= -(\mathbf{x} - \boldsymbol{\mu}_j)^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}_j) = -(\cancel{\mathbf{x}^{\mathrm{T}}\mathbf{x}} - 2\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\mu}_j)$$

$$\Rightarrow \; \mathbf{w}_j^{\mathrm{T}}\mathbf{x} + w_{j0}$$

$w_j = \mu \qquad w_{j0} = -\frac{1}{2}\|\mu_j\|^2$

Broj parametara?

$K \cdot n \Rightarrow O(n)$

Koji model od navedenih pet odabrati?

$\Rightarrow$ Ovisi o podatcima! Razlika je u složenosti!

Danas. . .

1. Jednodimenzijski Bayesov klasifikator za kontinuirane ulaze
2. Višedimenzijski Bayesov klasifikator za kontinuirane ulaze
3. Napomene

Bayesov klasifikator: komponente algoritma

(1) Model $\mathcal{H}$:

$$
\begin{aligned}
\mathcal{H} &= \{h(\mathbf{x}|\boldsymbol{\theta})\}_{\boldsymbol{\theta}} \\
h(\mathbf{x}|\boldsymbol{\theta}) &= \big(h_1(\mathbf{x}|\boldsymbol{\theta}_1), \dots, h_K(\mathbf{x}|\boldsymbol{\theta}_K)\big) \\
\boldsymbol{\theta} &= (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K) \\
h_j(\mathbf{x}|\boldsymbol{\theta}_j) &= \ln p(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) + \ln P(\mathcal{C}_j) \\
\boldsymbol{\theta}_j &= (\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, P(\mathcal{C}_j))
\end{aligned}
$$

generativan – diskriminativan ?

parametarski – neparametarski ?

linearan – nelinearan ?

(MLE)

pristanost ograničenjem – pristranost preferencijom ?

Bayesov klasifikator: komponente algoritma

(3) Optimizacijski postupak:

$$\boldsymbol{\theta}^* = \operatorname*{argmin}_{\boldsymbol{\theta}} E(\boldsymbol{\theta}|\mathcal{D}) = \operatorname*{argmin}_{\boldsymbol{\theta}} \mathbb{E}_{\mathcal{D}}[L]$$

(empirijska pogreška) (očekivanje f-je gubitka)

MLE:

$$\boldsymbol{\theta}^* = \operatorname*{argmax}_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta}) = \operatorname*{argmin}_{\boldsymbol{\theta}} \big(-p(\mathcal{D}|\boldsymbol{\theta})\big) = \operatorname*{argmin}_{\boldsymbol{\theta}} \big(-\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D})\big)$$

$$\Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta}^*|\mathcal{D}) = 0$$

(2) Funkcija gubitka L:

$$\mathbb{E}_{\mathcal{D}}[L] = \sum_{i=1}^{N} \frac{1}{N} L\big(y^{(i)}, h(\mathbf{x}^{(i)}|\boldsymbol{\theta})\big)$$

=

$$-\ln \mathcal{L}(\theta|\mathcal{D}) = -\ln \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta}) = \sum_{i=1}^{N} -\ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta})$$

Linearnost naivnog Bayesovog klasifikatora

Naivan Bayesov klasifikator (diskretni i kontinuirani) je linearan model!

Za kontinuirani NB smo pokazali da je linearan. Za diskretan NB:

Npr.

za $K=2$

$$P(\mathbf{x}|\mathcal{C}_j) \propto \prod_{i=1}^{n} P(x_i|\mathcal{C}_j) = \prod_{i=1}^{n} \underbrace{\mu_{i,j}^{x_i}(1-\mu_{i,j})^{1-x_i}}_{\text{Bernoulli}}$$

$$\begin{aligned} h_j(\mathbf{x}) &= \ln P(\mathcal{C}_j|\mathbf{x}) + \ln P(\mathcal{C}_j) \\ &= \sum_i \Big(x_i \ln \mu_{i,j} + (1-x_i)\ln(1-\mu_{i,j}) \Big) + \ln P(\mathcal{C}_j) \\ &= \sum_i \underbrace{\frac{\mu_{i,j}}{1-\mu_{i,j}}}_{w_i} x_i + \text{konst.} = \mathbf{w}_j^{\mathrm{T}}\mathbf{x} + w_{j0} \end{aligned}$$

$\Rightarrow \quad VC(\mathcal{H}) = n+1$

Složeniji modeli (polunaivni NB, model s nedijeljenom $\mathbf{\Sigma}$) su nelinearni i imaju $VC(\mathcal{H}) > n+1$.

Diskretizacija značajki

Kombinacija diskretnih (nominalnih!) i kontinuiranih značajki je problematična:

$$P(\mathcal{C}_j|\overbrace{x_1,\dots,x_k}^{\text{diskretne}},\overbrace{x_{k+1},\dots,x_n}^{\text{kontinuirane}}) \propto$$

$$P(\mathcal{C}_j)\underbrace{\prod_{i=1}^{k} P(x_i|\mathcal{C}_j)}_{\in[0,1]}\ \underbrace{\prod_{i=k+1}^{n} p(x_i|\mathcal{C}_j)}_{\in[0,\infty)}$$

nesumjerljivo!

Treba napraviti diskretizaciju kontinuiranih značajki:

- nadzirana vs. nenadzirana
  - nadzirana: *recursive minimal entropy partitioning (RMEP)*
  - nenadzirana: *equal width binning*, *equal frequency binning*
- globalna vs. lokalna
  - lokalna: grupiranje k-srednjih vrijednosti

Sažetak

- Za kontinuirane značajke izglednost klase modelira se **Gaussovom gustoćom** (modelira šum)
- Kod višedimenzijskog Bayesovog klasifikator izglednost je modelirana **multivarijatnom Gaussovom gustoćom**, čija kovarijacijska matrica mora biti **pozitivno definitna**
- Uvođenjem dodatnih pretpostavki (dijeljena/dijagonalna/izotropna kovarijacijska matrica) moguće je **pojednostaviti model** (smanjiti broj parametara)
- Diskretan/kontinuiran naivan Bayesov klasifikator je **linearan model**
- Ako imamo diskretne i kontinuirane značajke, ove potonje treba **diskretizirati**

Sljedeća tema: Regresija