13. Procjena parametara

Strojno učenje 1, UNIZG FER, ak. god. 2021./2022.

Jan Šnajder, natuknice s predavanja, v1.3

1 Motivacija

- **Probabilistički modeli** modeliraju vjerojatnosnu razdiobu primjera $\mathbf{x}$ i/ili oznaka y
- **Prednosti**: (1) temeljeni na teoriji vjerojatnosti, (2) vjerojatnosna predikcija, (3) ugradnja apriornog znanja, (4) prikladni za male skupove podataka
- Npr. **Bayesov klasifikator** – $P(y|\mathbf{x}) \propto P(\mathbf{x}|y)P(y)$
  - odabrati prikladne razdiobe za $P(\mathbf{x}|y)$ i $P(y)$
  - **procijeniti parametre** razdioba na temelju podataka $\Leftrightarrow$ učenje modela

2 Slučajne varijable

- X – slučajna varijabla (s.v.) sa skupom mogućih vrijednosti $\{x_i\}$
- **Diskretna s.v.**:
  - $P(X = x)$, kraće $P(x)$ – **vjerojatnost** da diskretna s.v. poprimi vrijednost x
  - $P(x_i) \geqslant 0$, $\sum_i P(x_i) = 1 \Rightarrow$ **diskretna razdioba (distribucija) vjerojatnosti**
- **Kontinuirana s.v.**:
  - $p(x)$ – **funkcija gustoće vjerojatnosti (PDF)**
  - $p(x) \geqslant 0$, $\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 \Rightarrow$ **kontinuirana razdioba (distribucija) vjerojatnosti**
- **Očekivanje** – prosječna vrijednost: $\mathbb{E}[X] = \sum_x xP(x)$ odnosno $\mathbb{E}[X] = \int_{-\infty}^{\infty} x\,p(x)\,\mathrm{d}x$
- **Varijanca** – očekivano odstupanje od očekivanja: $\mathrm{Var}(X) = \sigma_X^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- **Kovarijanca** – zajednička varijabilnost dviju varijabli:

  $$\mathrm{Cov}(X, Y) = \sigma_{X,Y} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

  $\Rightarrow \mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X,)$, $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- **Pearsonov koeficijent korelacije**: $\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \in [-1, +1]$

- $\rho_{X,Y} = +1 \Leftrightarrow$ pozitivna **linearna zavisnost**
- $\rho_{X,Y} = 0 \Leftrightarrow$ **linearna nezavisnost**
- $\rho_{X,Y} = -1 \Leftrightarrow$ negativna **linearna zavisnost**

$\Rightarrow$ ne mjeri **nelinearnu zavisnost**!

- **Matrica kovarijacije** – kovarijacija svih parova varijabli **slučajnog vektora** $(X_1, \ldots, X_n)$:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \ldots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \ldots & \mathrm{Var}(X_n) \end{pmatrix} = \mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^{\mathrm{T}}\Big]$$

  - simetrična i pozitivno semidefinitna
  - singularna (tj. nema inverz) ako postoje linearno zavisni retci ili ako $\sigma_i^2 = 0$
  - $\mathrm{Cov}(X_i, X_j) = 0 \Rightarrow$ **dijagonalna** kovarijacijska matrica, $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i^2)$
  - $\sigma_i^2 = \sigma \Rightarrow$ **izotropna** kovarijacijska matrica, $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$

- **Nezavisne varijable** $\Leftrightarrow P(X, Y) = P(X)P(Y)$
  - nezavisne varijable su nekorelirane, $\mathrm{Cov}(X, Y) = \rho_{X,Y} = 0$, ali obrat ne vrijedi!

3 Osnovne vjerojatnosne distribucije

- Diskretna varijabla:
  - Jednodimenzijska binarna: **Bernoullijeva razdioba**
  - Jednodimenzijska viševrijednosna: **kategorička (multinulijeva) razdioba**
  - Višedimenzijska: Konkatenirani vektor binarnih/viševrijednosnih varijabli
- Kontinuirana varijable:
  - Jednodimenzijska: **univarijatna normalna (Gaussova) razdioba**
  - Višedimenzijska: **multivarijatna normalna (Gaussova) razdioba**
- **Bernoullijeva razdioba** – binarna s.v.:

$$P(x|\mu) = \mu^x(1-\mu)^{1-x}$$

$\Rightarrow \mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \mu(1-\mu)$

- **Kategorička (multinulijeva) razdioba** – viševrijednosna diskretna s.v.:

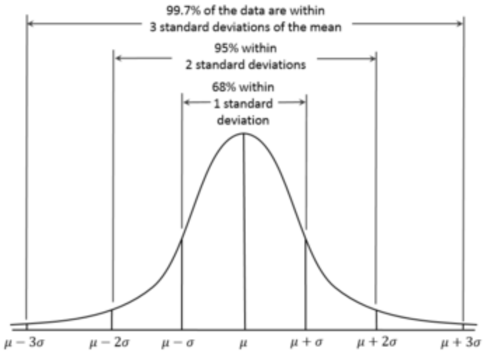
$$P(\mathbf{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

  - $\mathbf{x} = (x_1, x_2, \ldots, x_K)^{\mathrm{T}}$ – vektor indikatorskih varijabli (**one-hot encoding**)
  - $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$ – vjerojatnosti pojedinih vrijednosti, $\sum_k \mu_k = 1$, $\mu_k \geqslant 0$

- **Normalna (Gaussova) razdioba** – kontinuirana vrijednost uz prisustvo **šuma**:

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

  $\Rightarrow \mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$

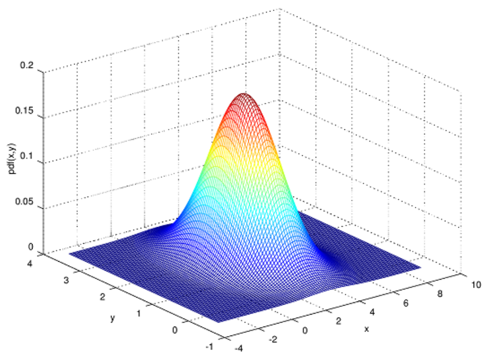


- **Multivarijatna (višedimenzijska) normalna (Gaussova) razdioba**:

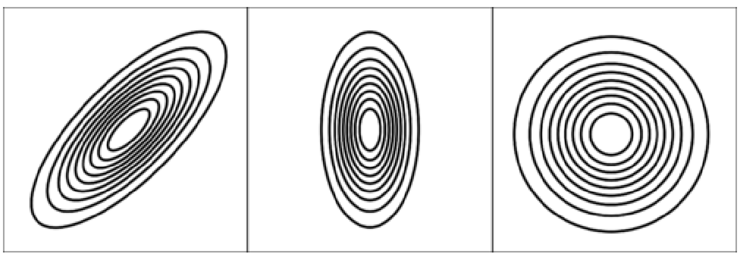
$$p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big)$$

  $\Rightarrow \mathbb{E}[X] = \boldsymbol{\mu}$, $\mathrm{Cov}(X_i, X_j) = \boldsymbol{\Sigma}_{ij}$

  - značajke su savršeno **multikolinearne** $\Leftrightarrow$ $\boldsymbol{\Sigma}$ je singularna $\Leftrightarrow$ $p(\mathbf{x})$ je nedefinirana
  - $\Delta^2 = (\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ – **kvadratna forma**
  - Δ – **Mahalanobisova udaljenost** između $\mathbf{x}$ i $\boldsymbol{\mu}$



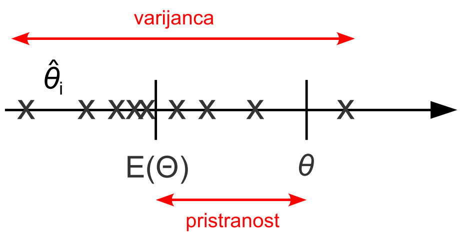
- Kovarijacijska matrica određuje izgled Gaussove razdiobe:



$\boldsymbol{\Sigma}$ $\qquad$ $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i^2)$ $\qquad$ $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$

4 Procjena parametara

- Raspolažemo konačnim i slučajnim (= reprezentativnim) uzorkom iz **populacije**
- Na temelju uzorka **procjenjujemo parametre** modela koji opisuje populaciju
- **Uzorak** – niz s.v. $(X_1, X_2, \dots, X_N)$ koje su **iid** (identično i nezavisno distribuirane)
- **Statistika** – funkcija slučajnog uzorka, $\Theta = g(X_1, X_2, \dots, X_N)$
- **Procjenitelj (estimator)** – statistika koja odgovara parametru populacije θ
- **Procjena (estimacija)** – vrijednost procjenitelja za dani uzorak, $\hat{\theta} = g(x_1, x_2, \dots, x_N)$
- Procjenitelj je s.v., pa ima svoje očekivanje i varijancu



- **Pristranost (bias)** – razlika između očekivanja procjenitelja i parametra populacije:

$$b(\Theta) = \mathbb{E}[\Theta] - \theta$$

- **Nepristran procjenitelj (unbiased estimator)** $\Leftrightarrow \mathbb{E}[\Theta] = \theta \Leftrightarrow b(\Theta) = 0$
- Primjeri:
  - $\hat{\mu} = \frac{1}{N} \sum_i x^{(i)}$ – nepristran procjenitelj srednje vrijednosti μ
  - $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \hat{\mu})^2$ – pristran procjenitelj varijance σ^2 (podcjenjuje!)
  - $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x^{(i)} - \hat{\mu})^2$ – nepristran procjenitelj varijance σ^2
- Postupci za izvođenje procjenitelja:
  - **Procjenitelj najveće izglednosti** (maximum likelihood estimator, **MLE**)
  - **Procjenitelj maximum aposteriori (MAP)**
  - **Bayesovski procjenitelj** (bayesian estimator)
- Radit ćemo MLE i MAP