Bayesova teorija odlučivanja

- Teorijski osnovi klasifikacije
  - LR test
  - Verovatnoća greške
  - Bayesov rizik
  - Bayesov, MAP i ML kriterijum
  - Odlučivanje kada postoji više klasa
  - Diskriminantne funkcije

Bayesova teorema u problemu klasifikacije

- Kod problema klasifikacije, Bayesova teorema se može izraziti kao:

$$\mathrm{P}(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}{\sum_{j=1}^{K} p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}$$

pri čemu je ω_j j-ta klasa, a $\mathbf{x}$ vektor obeležja

- $\mathrm{P}(\omega_j)$ Apriorna verovatnoća klase ω_j
- $\mathrm{P}(\omega_j|\mathbf{x})$ Aposteriorna verovatnoća klase ω_j ako je data opservacija $\mathbf{x}$
- $p(\mathbf{x}|\omega_j)$ Izglednost/verodostojnost opservacije $\mathbf{x}$ u klasi ω_j
- $p(\mathbf{x})$ Normalizaciona konstanta

Bayesova teorema u problemu klasifikacije

- Kod problema klasifikacije, Bayesova teorema se može izraziti kao:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{\sum_{j=1}^{K} p(\mathbf{x}|\omega_j)P(\omega_j)}$$

pri čemu je ω_j j-ta klasa, a **x** vektor obeležja

- $P(\omega_j)$ **Apriorna verovatnoća klase $\boldsymbol{\omega_j}$**
- $P(\omega_j|\mathbf{x})$ Aposteriorna verovatnoća klase ω_j ako je data opservacija **x**
- $p(\mathbf{x}|\omega_j)$ Izglednost/verodostojnost opservacije **x** u klasi ω_j
- $p(\mathbf{x})$ Normalizaciona konstanta

- **Apriorna verovatnoća klase $\boldsymbol{\omega_j}$**
  - Verovatnoća da uzorak pripada klasi ω_j bez obzira na to kako izgleda (bez obzira na **x**)
  - Odslikava naše prethodno znanje o problemu
    - S kojom verovatnoćom se na pokretnoj traci može naći losos ili brancin zavisi od stanja lovišta, sezone...

Bayesova teorema u problemu klasifikacije

- Kod problema klasifikacije, Bayesova teorema se može izraziti kao:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{\sum_{j=1}^{K} p(\mathbf{x}|\omega_j)P(\omega_j)}$$

pri čemu je ω_j j-ta klasa, a **x** vektor obeležja

- $P(\omega_j)$ Apriorna verovatnoća klase ω_j
- $P(\omega_j|\mathbf{x})$ **Aposteriorna verovatnoća klase ω_j** ako je data opservacija **x**
- $p(\mathbf{x}|\omega_j)$ Izglednost/verodostojnost opservacije **x** u klasi ω_j
- $p(\mathbf{x})$ Normalizaciona konstanta

- **Aposteriorna verovatnoća klase ω_j**
  - Verovatnoća da uzorak pripada klasi ω_j imajući u vidu to kako izgleda (imajući u vidu **x**)
  - Nakon što se registruje (izmeri) **x**, verovatnoća pripadnosti klasi ω_j više nije $P(\omega_j)$ nego $P(\omega_j|\mathbf{x})$

Bayesova teorema u problemu klasifikacije

- Kod problema klasifikacije, Bayesova teorema se može izraziti kao:

$$\mathrm{P}(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}{\sum_{j=1}^{K} p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}$$

  pri čemu je ω_j *j*-ta klasa, a **x** vektor obeležja
  - $\mathrm{P}(\omega_j)$ Apriorna verovatnoća klase ω_j
  - $\mathrm{P}(\omega_j|\mathbf{x})$ Aposteriorna verovatnoća klase ω_j ako je data opservacija **x**
  - $p(\mathbf{x}|\omega_j)$ Izglednost/verodostojnost opservacije **x** u klasi ω_j
  - $p(\mathbf{x})$ Normalizaciona konstanta
- Tipično pravilo odlučivanja kako dodeliti vektor obeležja klasi jeste **da se izabere klasa $\boldsymbol{\omega_j}$ sa najvećom verovatnoćom** $\mathrm{P}(\omega_j|\mathbf{x})$:

$$\hat{\omega} = \underset{\omega_j,\, j=1,2,\ldots K}{\arg\max}\, \mathrm{P}(\omega_j|\mathbf{x}) = \underset{\omega_j,\, j=1,2,\ldots K}{\arg\max} \frac{p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}{p(\mathbf{x})} = \underset{\omega_j,\, j=1,2,\ldots K}{\arg\max}\, p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)$$

  - Na ovaj način, intuitivno biramo klasu koja ima najveće izglede da je iz nje potekao dati vektor obeležja **x**

Binarna klasifikacija

- Slučaj kada imamo samo dve klase (binarna klasifikacija)

$$P(\omega_1|\mathbf{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2|\mathbf{x})$$

- Na osnovu Bayesovog pravila to se svodi na:

$$\frac{p(\mathbf{x}|\omega_1)P(\omega_1)}{p(\mathbf{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{p(\mathbf{x}|\omega_2)P(\omega_2)}{p(\mathbf{x})}$$

- Pošto $p(\mathbf{x})$ ne utiče na pravilo odlučivanja, može se eliminisati, i izraz se može napisati u obliku:

$$\boxed{\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)}}$$

  - Količnik $\Lambda(\mathbf{x})$ se naziva količnik izglednosti (eng. *likelihood ratio*), a odgovarajuće pravilo odlučivanja se naziva *LR test* (test količnika izglednosti)
  - Pretpostavka je da su raspodele verovatnoće obeležja u pojedinim klasama i apriorne verovatnoće klasa **poznate**
    - Ako već nisu unapred date, treba ih estimirati na osnovu podataka za obuku

LR test (primer 1)

- Formulisati pravilo odlučivanja prema količniku izglednosti za dve klase čije su apriorne verovatnoće jednake, a gustine raspodele verovatnoća su date izrazima:

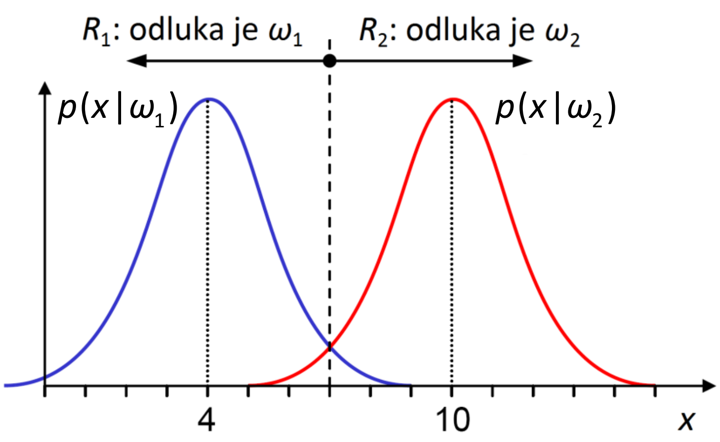
$$p(x\,|\,\omega_1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-4)^2}{2}},\ \ p(x\,|\,\omega_2)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-10)^2}{2}}.$$

- Rešenje:

$$\Lambda(x)=\frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-4)^2}{2}}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-10)^2}{2}}}=e^{-\frac{(x-4)^2}{2}+\frac{(x-10)^2}{2}}\ \underset{\omega_2}{\overset{\omega_1}{\gtrless}}\ 1$$

$$(x-10)^2-(x-4)^2\ \underset{\omega_2}{\overset{\omega_1}{\gtrless}}\ 0$$

$$x\ \underset{\omega_2}{\overset{\omega_1}{\lessgtr}}\ 7$$



LR test (primer 2)

- Formulisati pravilo odlučivanja prema količniku izglednosti za dve klase za čije apriorne verovatnoće važi P(ω_1) = 2P(ω_2), a gustine raspodele verovatnoća su date izrazima:

$$p(x\,|\,\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-4)^2}{2}},\ \ p(x\,|\,\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-10)^2}{2}}.$$

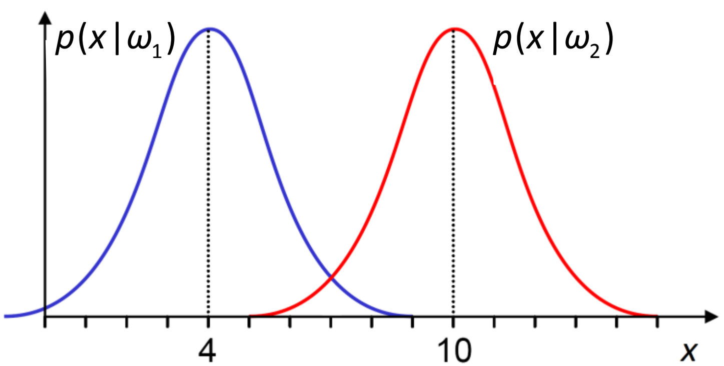
- Rešenje:

$$\Lambda(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-4)^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-10)^2}{2}}} = e^{-\frac{(x-4)^2}{2}+\frac{(x-10)^2}{2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{1}{2}$$

$$(x-10)^2 - (x-4)^2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 2\ln\frac{1}{2}$$

$$x \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 7 + \frac{\ln 2}{6} \approx 7{,}12$$

Granica odlučivanja pomera se ka centru raspodele sa manjom apriornom verovatnoćom



Verovatnoća greške

- Bez obzira na optimalno pravilo odlučivanja, greške su neizbežne
  - Verovatnoća greške opisuje kvalitet performansi pravila odlučivanja
- Na osnovu teoreme totalne verovatnoće, verovatnoća greške jednaka je:

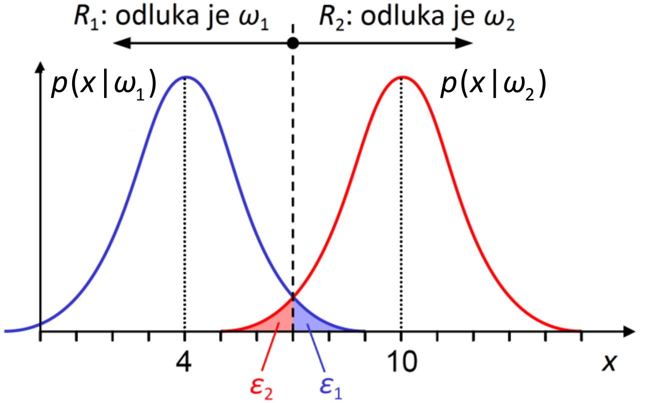
$$\mathrm{P}(\text{greška}) = \sum_{i=1}^{K} \mathrm{P}(\text{greška}|\,\omega_i)\,\mathrm{P}(\omega_i),$$

pri čemu je verovatnoća greške unutar određene klase jednaka:

$$\mathrm{P}(\text{greška}|\,\omega_i) = \mathrm{P}(\text{odabrano}\,\omega_j, j \neq i\,|\,\omega_i) = \int_{R_j, j\neq i} p(\mathbf{x}\,|\,\omega_i)\,\mathrm{d}\mathbf{x}$$

  - U najjednostavnijem slučaju, kada postoje dve klase, verovatnoća greške je:

$$\begin{aligned}\mathrm{P}(\text{greška}) &= \mathrm{P}(\omega_1)\mathrm{P}(\text{greška}\,|\,\omega_1) + \mathrm{P}(\omega_2)\mathrm{P}(\text{greška}\,|\,\omega_2)\\ &= \mathrm{P}(\omega_1)\underbrace{\int_{R_2} p(\mathbf{x}\,|\,\omega_1)\,\mathrm{d}\mathbf{x}}_{\varepsilon_1} + \mathrm{P}(\omega_2)\underbrace{\int_{R_1} p(\mathbf{x}\,|\,\omega_2)\,\mathrm{d}\mathbf{x}}_{\varepsilon_2}\end{aligned}$$



Bayesov rizik

- U opštem slučaju cena pogrešne klasifikacije uzorka iz klase ω_1 u klasu ω_2 nije ista kao kada se uzorak iz klase ω_2 pogrešno klasifikuje u klasu ω_1
  - Slučaj kada se bolestan čovek proglasi za zdravog opasniji je nego kada se zdrav čovek proglasi za bolesnog
- Ovaj koncept se može formalizovati pomoću *funkcije cene* C_{ij}
  - C_{ij} predstavlja cenu izbora klase ω_i ako je stvarna klasa ω_j (za $i = j$ obično $C_{ij} = 0$)
- *Bayesov rizik* definiše se kao matematičko očekivanje cene:

$$\begin{aligned}\Re = E[C] &= \sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\text{odabrano } \omega_i \text{ a stvarna klasa je } \omega_j) \\ &= \sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\mathbf{x} \in R_i \mid \omega_j)\,\mathrm{P}(\omega_j)\end{aligned}$$

pri čemu:

$$\mathrm{P}(\mathbf{x} \in R_i \mid \omega_j) = \int_{R_i} p(\mathbf{x} \mid \omega_j)\,\mathrm{d}\mathbf{x}$$

- Kako izgleda pravilo odlučivanja koje minimizuje Bayesov rizik?

Bayesov rizik (slučaj dve klase)

$$
\begin{aligned}
\Re = E[C] &= \sum_{i=1}^{2}\sum_{j=1}^{2} C_{ij}\cdot \mathrm{P}(\text{odabrano } \omega_i \text{ a stvarna klasa je } \omega_j) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} C_{ij}\cdot \mathrm{P}(\mathbf{x}\in R_i \mid \omega_j)\,\mathrm{P}(\omega_j) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} C_{ij}\cdot \mathrm{P}(\omega_j)\int_{R_i} p(\mathbf{x}\mid\omega_j)\,\mathrm{d}\mathbf{x} \\
&= C_{11}\cdot \mathrm{P}(\omega_1)\int_{R_1} p(\mathbf{x}\mid\omega_1)\,\mathrm{d}\mathbf{x} + C_{12}\cdot \mathrm{P}(\omega_2)\int_{R_1} p(\mathbf{x}\mid\omega_2)\,\mathrm{d}\mathbf{x} \\
&\quad + C_{21}\cdot \mathrm{P}(\omega_1)\int_{R_2} p(\mathbf{x}\mid\omega_1)\,\mathrm{d}\mathbf{x} + C_{22}\cdot \mathrm{P}(\omega_2)\int_{R_2} p(\mathbf{x}\mid\omega_2)\,\mathrm{d}\mathbf{x} \\
&= C_{11}\cdot \mathrm{P}(\omega_1)\int_{R_1} p(\mathbf{x}\mid\omega_1)\,\mathrm{d}\mathbf{x} + C_{12}\cdot \mathrm{P}(\omega_2)\int_{R_1} p(\mathbf{x}\mid\omega_2)\,\mathrm{d}\mathbf{x} \\
&\quad + C_{21}\cdot \mathrm{P}(\omega_1)\Big(1-\int_{R_1} p(\mathbf{x}\mid\omega_1)\,\mathrm{d}\mathbf{x}\Big) + C_{22}\cdot \mathrm{P}(\omega_2)\Big(1-\int_{R_1} p(\mathbf{x}\mid\omega_2)\,\mathrm{d}\mathbf{x}\Big) \\
&= C_{21}\cdot \mathrm{P}(\omega_1) + C_{22}\cdot \mathrm{P}(\omega_2) \\
&\quad + (C_{12}-C_{22})\,\mathrm{P}(\omega_2)\int_{R_1} p(\mathbf{x}\mid\omega_2)\,\mathrm{d}\mathbf{x} - (C_{21}-C_{11})\,\mathrm{P}(\omega_1)\int_{R_1} p(\mathbf{x}\mid\omega_1)\,\mathrm{d}\mathbf{x}
\end{aligned}
$$

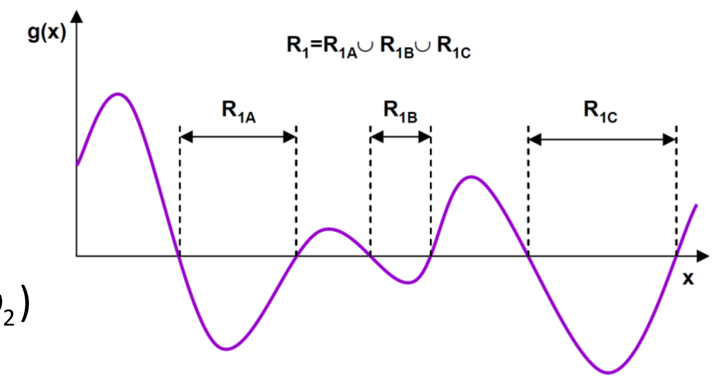
- Potrebno je odabrati region odlučivanja kojim se $\Re$ minimizuje, pri čemu je jasno da sabirci $C_{21}\cdot\mathrm{P}(\omega_1)$ i $C_{22}\cdot\mathrm{P}(\omega_2)$ ne utiču na to

Bayesov rizik (slučaj dve klase)

$$\hat{R}_1 = \arg\min_{R_1} \{\Re\}$$

$$= \arg\min_{R_1} \left\{ \int_{R_1} \left[(C_{12} - C_{22}) \mathrm{P}(\omega_2)\, p(\mathbf{x} \mid \omega_2) - (C_{21} - C_{11}) \mathrm{P}(\omega_1)\, p(\mathbf{x} \mid \omega_1) \right] \mathrm{d}\mathbf{x} \right\}$$

$$= \arg\min_{R_1} \left\{ \int_{R_1} g(\mathbf{x}) \mathrm{d}\mathbf{x} \right\}$$



- Pošto je potrebno minimizovati integral, treba odabrati samo oblasti u kojima $g(\mathbf{x}) < 0$:

$$(C_{21} - C_{11}) \mathrm{P}(\omega_1)\, p(\mathbf{x} \mid \omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} (C_{12} - C_{22}) \mathrm{P}(\omega_2)\, p(\mathbf{x} \mid \omega_2)$$

što se ponovo svodi na određenu varijantu LR testa:

$$\boxed{\frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{(C_{12} - C_{22})}{(C_{21} - C_{11})} \cdot \frac{\mathrm{P}(\omega_2)}{\mathrm{P}(\omega_1)}}$$

- Dakle, minimizacija Bayesovog rizika takođe se svodi na primenu LR testa

Bayesov rizik (primer)

- Neka je potrebno izvršiti klasifikaciju u dve klase za koje su raspodele obeležja date sledećim izrazima:

$$p(\mathbf{x}|\omega_1)=\frac{1}{\sqrt{2\pi}\sqrt{3}}e^{-\frac{x^2}{6}},\quad p(\mathbf{x}|\omega_2)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}}.$$

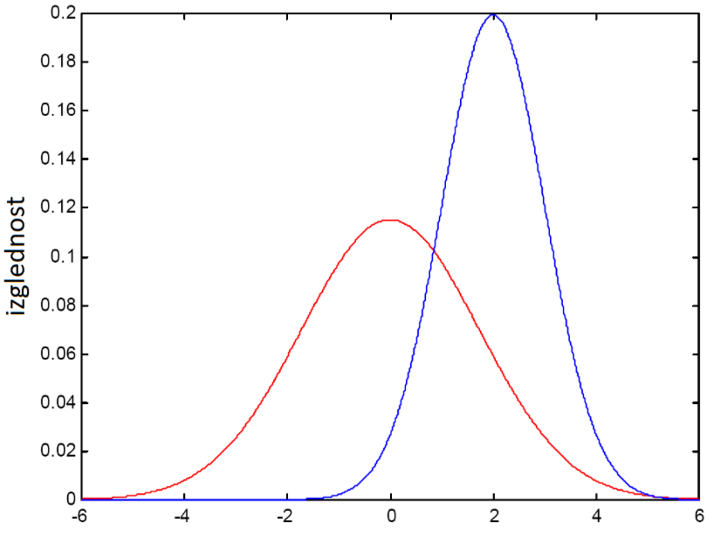
  Skicirati raspodele obeležja i, uz pretpostavku da je $P(\omega_1)=P(\omega_2)=0.5$, $C_{11}=C_{22}=0$, $C_{12}=1$ i $C_{21}=\sqrt{3}$, odrediti pravilo odlučivanja koje minimizuje Bayesov rizik.

- Rešenje:

$$\Lambda(x)=\frac{\frac{1}{\sqrt{2\pi}\sqrt{3}}e^{-\frac{x^2}{6}}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}}}=\frac{1}{\sqrt{3}}e^{-\frac{x^2}{6}+\frac{(x-2)^2}{2}}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}\frac{1}{\sqrt{3}}$$

$$-\frac{x^2}{6}+\frac{(x-2)^2}{2}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}0$$

$$\hat{R}_1=(-\infty,3-\sqrt{3}]\cup[3+\sqrt{3},\infty)$$
$$=(-\infty,1.27]\cup[4.73,\infty)$$



Bayesov rizik (primer)

- Neka je potrebno izvršiti klasifikaciju u dve klase za koje su raspodele obeležja date sledećim izrazima:

$$p(\mathbf{x}\,|\,\omega_1)=\frac{1}{\sqrt{2\pi}\sqrt{3}}e^{-\frac{x^2}{6}},\quad p(\mathbf{x}\,|\,\omega_2)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}}.$$

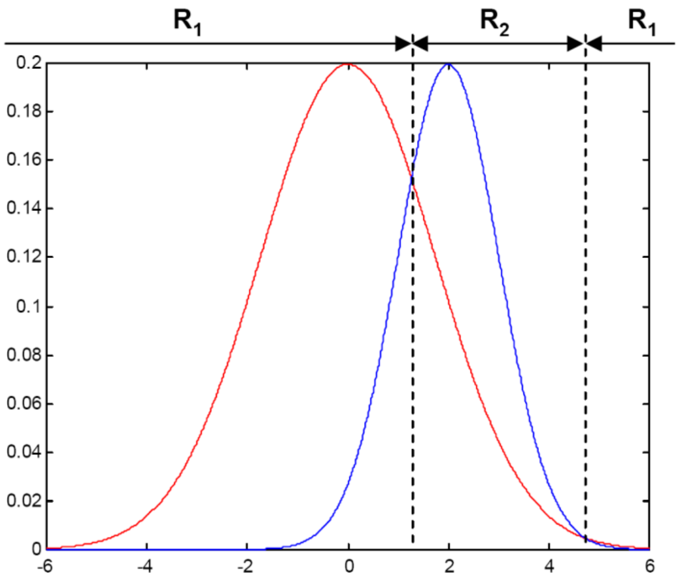
  Skicirati raspodele obeležja i, uz pretpostavku da je $P(\omega_1)=P(\omega_2)=0.5$, $C_{11}=C_{22}=0$, $C_{12}=1$ i $C_{21}=\sqrt{3}$, odrediti pravilo odlučivanja koje minimizuje Bayesov rizik.

- Rešenje:

$$\Lambda(x)=\frac{\frac{1}{\sqrt{2\pi}\sqrt{3}}e^{-\frac{x^2}{6}}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}}}=\frac{1}{\sqrt{3}}e^{-\frac{x^2}{6}+\frac{(x-2)^2}{2}}\ \underset{\omega_2}{\overset{\omega_1}{\gtrless}}\ \frac{1}{\sqrt{3}}$$

$$-\frac{x^2}{6}+\frac{(x-2)^2}{2}\ \underset{\omega_2}{\overset{\omega_1}{\gtrless}}\ 0$$

$$\hat{R}_1=(-\infty,3-\sqrt{3}]\cup[3+\sqrt{3},\infty)$$
$$=(-\infty,1.27]\cup[4.73,\infty)$$



Rezime

- Pravilo odlučivanja zasnovano na LR testu koje minimizuje Bayesov rizik često se naziva *Bayesov kriterijum*:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{(C_{12}-C_{22})}{(C_{21}-C_{11})} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Bayesov kriterijum

- Ako je cilj minimizovati verovatnoću greške, to je specijalan slučaj Bayesovog kriterijuma kada su cene simetrične i jednake 0 ili 1. Ova varijanta pravila odlučivanja zasnovanog na LR testu naziva se kriterijum *maksimalne aposteriorne verovatnoće* (MAP kriterijum)

$$C_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \qquad \Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)} \Leftrightarrow \frac{P(\omega_1|\mathbf{x})}{P(\omega_2|\mathbf{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 1$$

MAP kriterijum

- Ako su pri tome i apriorne verovatnoće jednake, $P(\omega_1) = P(\omega_2) = 1/2$, pravilo odlučivanja zasnovano na LR testu se svodi na *kriterijum maksimalne izglednosti* (ML kriterijum), koji maksimizuje izglednost $P(\mathbf{x}|\omega_i)$

$$C_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \wedge P(\omega_1) = P(\omega_2) = 1/2 \qquad \Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 1$$

ML kriterijum

Još neke varijante LR testa

- Neyman-Pearsonov kriterijum
  - Maksimalna verovatnoća greške za jednu klasu se fiksira na određenu vrednost, nakon čega se minimizuje verovatnoća greške za drugu klasu
    - Kod binarne klasifikacije obično se fiksira verovatnoća lažnog pozitivnog rezultata (npr. pri detekciji radarom), ali ima i drugih okolnosti kada je ovaj pristup pogodan
      - Npr. u problemu klasifikacije ribe, moguće je da postoji propis kojim se zabranjuje da više od 1% lososa bude klasifikovano kao brancin
    - NP kriterijum je veoma popularan pošto ne zahteva predznanje o apriornim verovatnoćama pojedinih klasa niti cenama
- *Minimax* kriterijum
  - Zasniva se na minimizaciji gornjeg ograničenja Bayesovog rizika (Bayesovog rizika u najgorem slučaju) i često se koristi u teoriji igara
    - *Minimax* kriterijum ne zahteva poznavanje apriornih verovatnoća pojedinih klasa, ali se cene moraju definisati
    - Pogodan u slučajevima kada apriorne verovatnoće mogu varirati

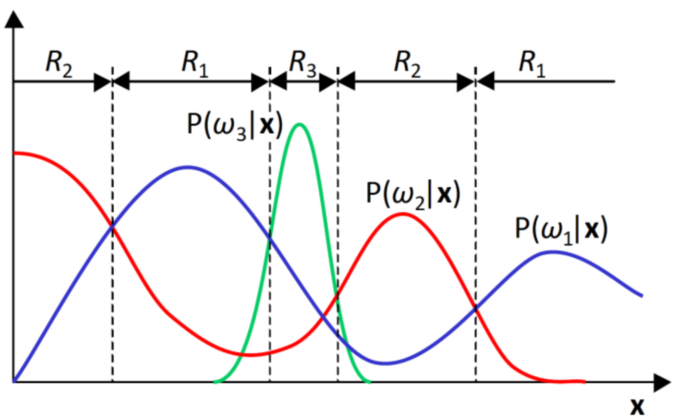
Minimizacija verovatnoće greške za više od dve klase

- Verovatnoća tačne klasifikacije iznosi:

$$\mathrm{P}(\text{tačno}) = 1 - \mathrm{P}(\text{greška}) = \sum_{i=1}^{K} \mathrm{P}(\omega_i) \int_{R_i} p(\mathbf{x} \mid \omega_i) \mathrm{d}\mathbf{x}$$

$$= \sum_{i=1}^{K} \int_{R_i} \mathrm{P}(\omega_i) p(\mathbf{x} \mid \omega_i) \mathrm{d}\mathbf{x} = \sum_{i=1}^{K} \underbrace{\int_{R_i} p(\mathbf{x}) \mathrm{P}(\omega_i \mid \mathbf{x}) \mathrm{d}\mathbf{x}}_{I_i}$$

  - Maksimizacija P(tačno) podrazumeva maksimizaciju svakog integrala I_i, tako da za svako **x** treba odabrati klasu za koju je $\mathrm{P}(\omega_i|\mathbf{x})$ maksimalno (odnosno, R_i treba definisati kao region u kom je $\mathrm{P}(\omega_i|\mathbf{x})$ maksimalno)
  - Dakle, i u slučaju više od dve klase, MAP kriterijum minimizuje verovatnoću greške



Minimizacija Bayesovog rizika za više od dve klase

- Pravilo odlučivanja koje minimizuje Bayesov rizik za dve klase može se jednostavno uopštiti i na slučaj sa tri ili više klasa

$$\begin{aligned}
\Re = E[C] &= \sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\text{odabrano } \omega_i \text{ a stvarna klasa je } \omega_j) \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\mathbf{x} \in R_i \,|\, \omega_j)\,\mathrm{P}(\omega_j) \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\omega_j) \int_{R_i} p(\mathbf{x} \,|\, \omega_j)\,\mathrm{d}\mathbf{x} \\
&= \sum_{i=1}^{K}\int_{R_i}\sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\omega_j)\, p(\mathbf{x} \,|\, \omega_j)\,\mathrm{d}\mathbf{x}
\end{aligned}$$

  - Minimizacija $\Re$ postiže se minimizacijom svakog integrala pojedinačno, što znači da R_i treba birati tako da bude:

$$\mathbf{x} \in R_i \;\Leftrightarrow\; \sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\omega_j)\, p(\mathbf{x} \,|\, \omega_j) < \sum_{j=1}^{K} C_{kj} \cdot \mathrm{P}(\omega_j)\, p(\mathbf{x} \,|\, \omega_j),\ \forall k \neq i$$

što se ponovo svodi na LR test, odnosno, uzorak **x** će se smestiti u klasu ω_i ako važi:

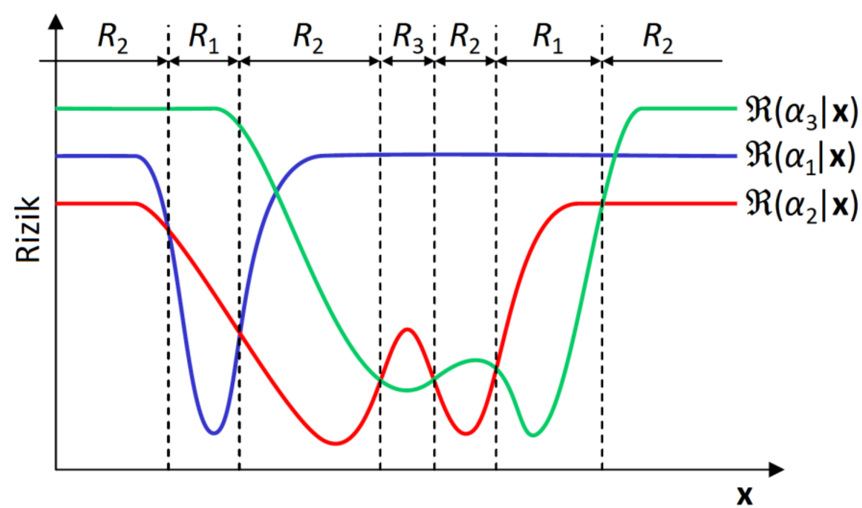
$$\frac{p(\mathbf{x} \,|\, \omega_i)}{p(\mathbf{x} \,|\, \omega_j)} > \frac{(C_{ij} - C_{jj})}{(C_{ji} - C_{ii})} \cdot \frac{\mathrm{P}(\omega_j)}{\mathrm{P}(\omega_i)},\quad \forall j \neq i$$

Minimizacija Bayesovog rizika za više od dve klase

- Moguća je i alternativna formulacija
  - Neka je α_i odluka da se uzorak **x** klasifikuje u klasu ω_i
  - Rizik (prosečna cena) te odluke jednak je:

$$\Re(\alpha_i \,|\, \mathbf{x}) = \sum_{j=1}^{K} C_{ij} \cdot \mathrm{P}(\omega_j \,|\, \mathbf{x})$$

  - Za svako **x** treba odabrati ono ω_i koje daje najmanje $\Re(\alpha_i|\mathbf{x})$, tj. minimalni rizik klasifikacije u ω_i



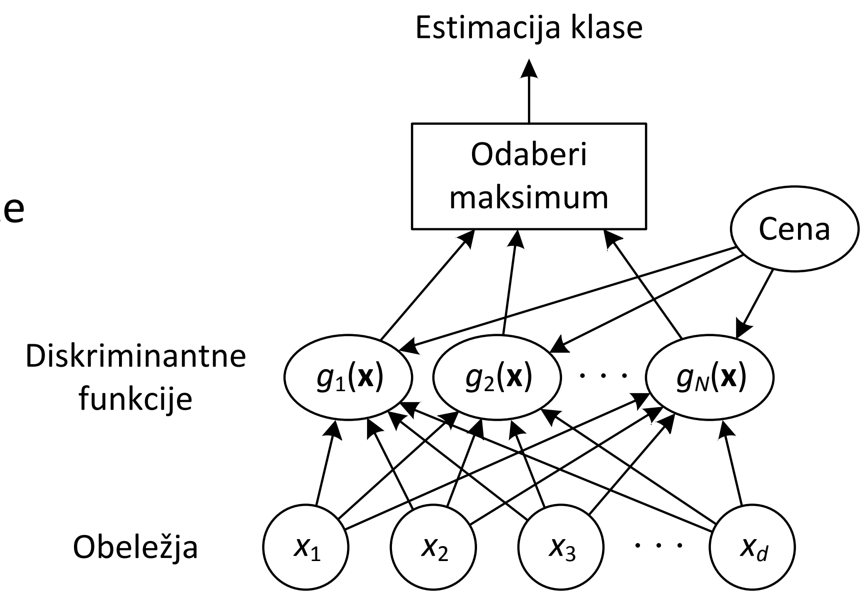
Diskriminantne funkcije

- Sva dosad prikazana pravila odlučivanja imaju istu strukturu
  - U svakoj tački **x** prostora obeležja bira se klasa ω_i koja maksimizuje (ili minimizuje) neku meru $g_i(\mathbf{x})$, koju nazivamo *diskriminantna funkcija*
  - Pravilo odlučivanja u tom slučaju glasi:

„dodeli **x** klasi ω_i ako je $g_i(\mathbf{x}) \geq g_j(\mathbf{x})$ za svako $j \neq i$"

odnosno, svodi se na poređenje diskriminanata za različite klase

  - U nekim slučajevima diskriminante se mogu svesti na jednostavne izraze
  - Ista monotona transformacija svih diskriminantnih funkcija ne utiče na rezultat, npr:
    - dodavanje konstante
    - logaritmovanje



Diskriminantne funkcije

- Tri osnovna pravila odlučivanja mogu se predstaviti preko diskriminantnih funkcija na sledeći način:

Kriterijum	Diskriminantna funkcija
Bayesov	$g_i(\mathbf{x}) = -\Re(\alpha_i \mid \mathbf{x})$
MAP	$g_i(\mathbf{x}) = \mathrm{P}(\omega_i \mid \mathbf{x})$
ML	$g_i(\mathbf{x}) = p(\mathbf{x} \mid \omega_i)$

- Bayesova teorija odlučivanja daje dobru teorijsku osnovu za klasifikaciju
  - U praksi apriorne verovatnoće pojedinih klasa, kao ni raspodele obeležja unutar tih klasa, **nisu poznate**, već ih je potrebno proceniti
  - Mnogi klasifikatori radije direktno određuju granice odlučivanja, koje su u opštem slučaju suboptimalne u odnosu na Bayesove klasifikatore
    - Linearne diskriminantne funkcije, perceptroni, vektori nosači, neuralne mreže...