



Uvod u umjetnu inteligenciju

Doc. dr. sc. Danijel Zelenika, danijel.zelenika@fsre.sum.ba

Naivni Bayesov klasifikator

Bayesovo pravilo

- Bayesovo pravilo (vjerojatnost da primjer \mathbf{x} pripada klasi c)

$$P(c | \mathbf{x}) = \frac{P(\mathbf{x} | c) P(c)}{P(\mathbf{x})}$$

- Gdje je:

- c – klasa (hipoteza)
- primjer $\mathbf{x} = (x_1, \dots, x_n)$ – skup atributa
- $P(c)$ – apriorna vjerojatnost klase (eng. prior)
- $P(\mathbf{x} | c)$ – je uvjetna vjerojatnost atributa \mathbf{x} , pod uvjetom da je klasa c (eng. class likelihood)
- $P(c | \mathbf{x})$ – je aposteriorna vjerojatnost klase (eng. posterior)
- $P(\mathbf{x})$ – ukupna vjerojatnost atributa

Naivni Bayes

- Ako pretpostavimo da su atributu međusobno neovisni unutar klase (što je „naivna” pretpostavka), uvjetna vjerojatnost $P(x | c)$ se razlaže na sljedeće:

$$P(x | c) = \prod_{i=1}^n P(x_i | c)$$

- $P(x)$ se izostavlja jer je isti za sve klase pa ne utječe na usporedbu aposteriornih vjerojatnosti ni na izbor najvjerojatnije klase.
- Za klasifikaciju ga možemo izostaviti, ali za računanje stvarnih vjerojatnost ne

Naivni Bayesov klasifikator

- Konačna formula:

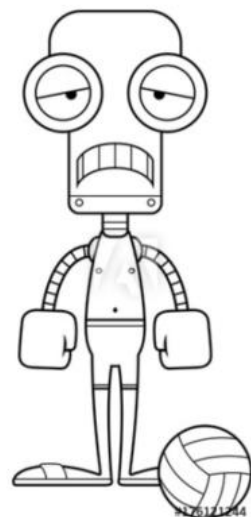
$$P(c|x) = P(c) \prod_{i=1}^n P(x_i|c)$$

- Za klasifikacijsku odluku se koristi hipoteza maksimum a posteriori ili MAP- hipoteza

$$h_{MAP} = \operatorname{argmax}(P(c) \prod_{i=1}^n P(x_i|c))$$

Primjer: Dan za odbojku na pijesku

	x_1	x_2	x_3	x_4	y
Dan	Vrijeme	Temp	Vlažnost	Vjetar	Odbojka?
1	sunčano	visoka	visoka	slab	ne
2	sunčano	visoka	visoka	jak	ne
3	oblačno	visoka	visoka	slab	da
4	kišno	srednja	visoka	slab	da
5	kišno	niska	normalna	slab	da
6	kišno	niska	normalna	jak	ne
7	oblačno	niska	normalna	jak	da
8	sunčano	srednja	visoka	slab	ne
9	sunčano	niska	normalna	slab	da
10	kišno	srednja	normalna	slab	da
11	sunčano	srednja	normalna	jak	da
12	oblačno	srednja	visoka	jak	da
13	oblačno	visoka	normalna	slab	da
14	kišno	srednja	visoka	jak	ne



Primjer: Dan za odbojku na pijesku

- Zanima nas klasifikacija za novi (petnaesti primjer):

$$\mathbf{x} = (\underbrace{\text{Vrijeme}=\text{sunčano}}_{x_1}, \underbrace{\text{Temp}=\text{niska}}_{x_2}, \underbrace{\text{Vlažnost}=\text{visoka}}_{x_3}, \underbrace{\text{Vjetar}=\text{jak}}_{x_4})$$

- MAP- hipoteza je:

$$h_{\text{MAP}} = \operatorname{argmax}_{y \in \{\text{da}, \text{ne}\}} P(y|\mathbf{x})$$

$$\stackrel{u.n.}{=} \operatorname{argmax}_{y \in \{\text{da}, \text{ne}\}} P(y) \prod_{j=1}^n P(x_j|y)$$

$$= \operatorname{argmax}_{y \in \{\text{da}, \text{ne}\}} P(y) P(\text{sunčano}|y) P(\text{niska}|y) P(\text{visoka}|y) P(\text{jak}|y)$$

Primjer: Dan za odbojku na pijesku

□ Dakle trebamo izračunati:

$$P(\text{odbojka}|\text{da}) = P(\text{da})P(\text{sunčano}|\text{da})P(\text{niska}|\text{da})P(\text{visoka}|\text{da})P(\text{jak}|\text{da})$$

$$P(\text{odbojka}|\text{ne}) = P(\text{ne})P(\text{sunčano}|\text{ne})P(\text{niska}|\text{ne})P(\text{visoka}|\text{ne})P(\text{jak}|\text{ne})$$

□ Kako bi procijenili $P(\text{da})$ i $P(\text{ne})$? Najjednostavnija je relativna frekvencija (udio primjera u skupu za učenje koji imaju oznaku $y = \text{da}$, odnosno oznaku $y = \text{ne}$)

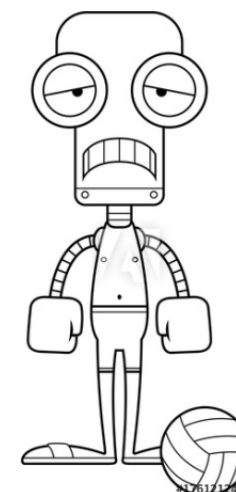
$$P(\text{da}) = \frac{9}{14} = 0.64$$

$$P(\text{ne}) = \frac{5}{14} = 0.36$$

Primjer: Dan za odbojku na pijesku

	x_1	x_2	x_3	x_4	y
Dan	Vrijeme	Temp	Vlažnost	Vjetar	Odbojka?
1	sunčano	visoka	visoka	slab	ne
2	sunčano	visoka	visoka	jak	ne
3	oblačno	visoka	visoka	slab	da
4	kišno	srednja	visoka	slab	da
5	kišno	niska	normalna	slab	da
6	kišno	niska	normalna	jak	ne
7	oblačno	niska	normalna	jak	da
8	sunčano	srednja	visoka	slab	ne
9	sunčano	niska	normalna	slab	da
10	kišno	srednja	normalna	slab	da
11	sunčano	srednja	normalna	jak	da
12	oblačno	srednja	visoka	jak	da
13	oblačno	visoka	normalna	slab	da
14	kišno	srednja	visoka	jak	ne

$$P(\text{da}) = \frac{9}{14}$$



Primjer: Dan za odbojku na pijesku

- Slično možemo procijeniti uvjetnu vjerojatnosti. Npr. $P(\text{sunčano}|\text{da})$ je udio primjera sa $x_1 = \text{sunčano}$ u podskupu primjera sa $y = \text{da}$

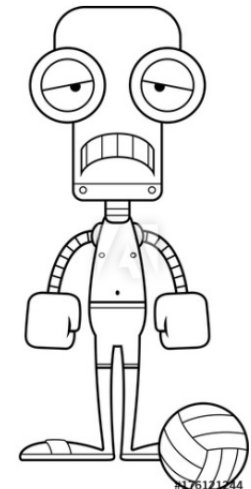
$$P(\text{sunčano}|\text{da}) = \frac{2}{9} = 0.22$$

$$P(\text{sunčano}|\text{ne}) = \frac{3}{5} = 0.6$$

Primjer: Dan za odbojku na pijesku

	x_1	x_2	x_3	x_4	y
Dan	Vrijeme	Temp	Vlažnost	Vjetar	Odbojka?
1	sunčano	visoka	visoka	slab	ne
2	sunčano	visoka	visoka	jak	ne
3	oblačno	visoka	visoka	slab	da
4	kišno	srednja	visoka	slab	da
5	kišno	niska	normalna	slab	da
6	kišno	niska	normalna	jak	ne
7	oblačno	niska	normalna	jak	da
8	sunčano	srednja	visoka	slab	ne
9	sunčano	niska	normalna	slab	da
10	kišno	srednja	normalna	slab	da
11	sunčano	srednja	normalna	jak	da
12	oblačno	srednja	visoka	jak	da
13	oblačno	visoka	normalna	slab	da
14	kišno	srednja	visoka	jak	ne

$$P(\text{sunčano}|\text{da}) \\ = \frac{2}{9}$$



Primjer: Dan za odbojku na pijesku

□ Izračun svih potrebnih vjerojatnosti:

$$P(y = \text{da}) = 9/14 = 0.64$$

$$P(y = \text{ne}) = 5/14 = 0.36$$

$$P(x_1 = \text{sunčano} | y = \text{da}) = 2/9 = 0.22$$

$$P(x_1 = \text{sunčano} | y = \text{ne}) = 3/5 = 0.6$$

$$P(x_2 = \text{niska} | y = \text{da}) = 3/9 = 0.33$$

$$P(x_2 = \text{niska} | y = \text{ne}) = 1/5 = 0.2$$

$$P(x_3 = \text{visoka} | y = \text{da}) = 3/9 = 0.33$$

$$P(x_3 = \text{visoka} | y = \text{ne}) = 4/5 = 0.8$$

$$P(x_4 = \text{jak} | y = \text{da}) = 3/9 = 0.33$$

$$P(x_4 = \text{jak} | y = \text{ne}) = 3/5 = 0.6$$

□ Izračun MAP-hipoteze

$$\begin{aligned} P(\text{odbojka} | \text{da}) &= P(\text{da}) P(\text{sunčano} | \text{da}) P(\text{niska} | \text{da}) P(\text{visoka} | \text{da}) P(\text{jak} | \text{da}) = \\ &0.64 \cdot 0.22 \cdot 0.33 \cdot 0.33 \cdot 0.33 = 0.0051 \end{aligned}$$

$$\begin{aligned} P(\text{odbojka} | \text{ne}) &= P(\text{ne}) P(\text{sunčano} | \text{ne}) P(\text{niska} | \text{ne}) P(\text{visoka} | \text{ne}) P(\text{jak} | \text{ne}) = \\ &0.36 \cdot 0.6 \cdot 0.2 \cdot 0.8 \cdot 0.6 = 0.0207 \end{aligned}$$

Primjer: Dan za odbojku na pijesku

□ Ako želimo, možemo izračunati i vjerojatnosti klasa:

$$P(\text{da}|\mathbf{x}) = \frac{0.0051}{0.0051 + 0.0207} = 0.198 \quad P(\text{ne}|\mathbf{x}) = \frac{0.0207}{0.0051 + 0.0207} = 0.802$$

Potrebno je implementirati naivni Bayesov klasifikator koji će, na temelju zadanog skupa podataka, procijeniti hoće li osoba igrati ili ne igrati odbojku na pijesku.

Program treba:

- učitati podatke,
- izračunati sve potrebne uvjetne vjerojatnosti,
- primijeniti Bayesovo pravilo,
- klasificirati novi podatak.

Napisati funkciju, npr. `predict(novi_podatak)` i vraća:

- „da„
- ili „ne"