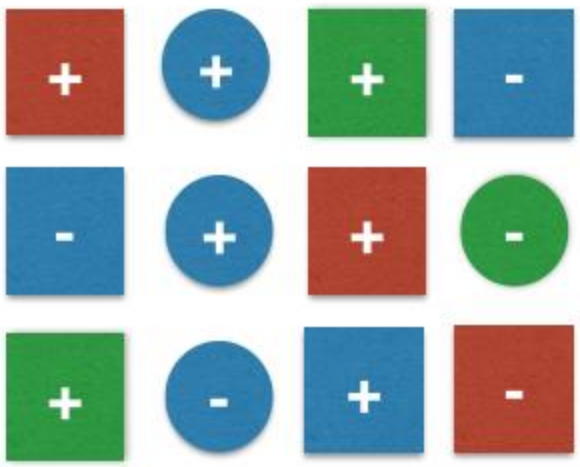
Multinominal Naïve Bayes – primer



Raschka, S., 2014. Naive bayes and text classification i-introduction and theory. *arXiv preprint arXiv:1410.5329*.

$y \in \{+, -\}, \quad x = [x_1, x_2], \quad N = 12$

$x_1 \in \{\text{b}, \text{g}, \text{r}, \text{y}\}, x_2 \in \{\text{c}, \text{s}\}$

Kako bismo klasifikovali ■?

Koristićemo ML estimaciju:

$$P(+) = \frac{7}{12} = 0.58 \qquad P(-) = \frac{5}{12} = 0.42$$

Naivna pretpostavka: boja i oblik su nezavisni

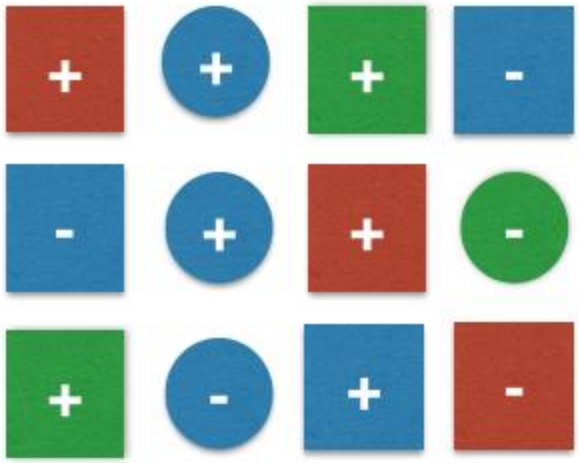
$$P(x = [\text{b}, \text{s}] | +) = P(\text{b}| +) \cdot P(\text{s}| +) = \frac{3}{7} \cdot \frac{5}{7} = 0.31$$

$$P(x = [\text{b}, \text{s}] | -) = P(\text{b}| -) \cdot P(\text{s}| -) = \frac{3}{5} \cdot \frac{3}{5} = 0.36$$

$$P(+|x = [\text{b}, \text{s}]) = P(x| +)P(+) = 0.31 \cdot 0.58 = 0.18$$

$$P(-|x = [\text{b}, \text{s}]) = P(x| -)P(-) = 0.36 \cdot 0.42 = 0.15$$

Multinominal Naïve Bayes – primer



Kako bismo klasifikovali ■ ?

$$P(x = [\mathrm{y}, \mathrm{s}]|+) = P(\mathrm{y}|+) \cdot P(\mathrm{s}|+) = 0 \cdot \frac{5}{7} = 0$$

$$P(x = [\mathrm{y}, \mathrm{s}]|-) = P(\mathrm{y}|-) \cdot P(\mathrm{s}|-) = 0 \cdot \frac{3}{5} = 0$$

$$P(+|x = [\mathrm{y}, \mathrm{s}]) = P(x = [y, s]|+)P(+)$$
$$= 0 \cdot 0.58 = 0$$
$$P(-|x = [\mathrm{y}, \mathrm{s}]) = P(x = [y, s]|-)P(-)$$
$$= 0 \cdot 0.42 = 0$$

- Odsustvo svega jedne vrednosti jednog atributa čini da ceo slog ima verovatnoću 0!

Smoothing

- Kako bismo izbegli verovatnoće 0 usled nedostatka reprezentativnih primera, dodaćemo *smoothing term* u multinominalni NB:

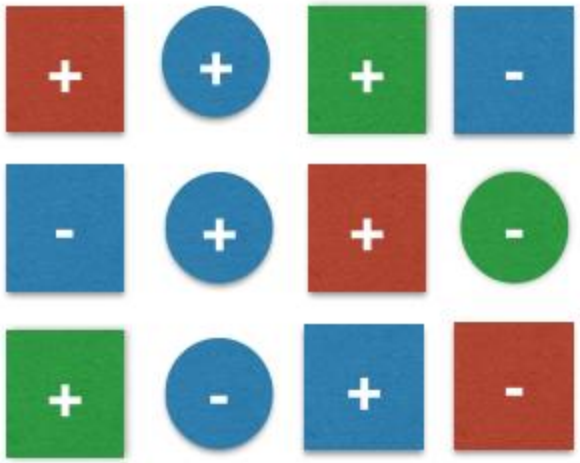
$$P(x_d|y = c) = \frac{N_{x_d,c} + \alpha}{N_c + \alpha V}$$

α – *parameter for additive smoothing*

V – broj mogućih vrednosti obeležja d

- Dodavanje αV „fiktivnih“ primera jednako raspoređenih po svim mogućim vrednostima obeležja x_d
- $\alpha < 1 \rightarrow$ Lidstone smoothing
- $\alpha = 1 \rightarrow$ Laplace smoothing

Primer sa Laplace smoothing



Kako bismo klasifikovali ■ ?

$$P(x = [\mathrm{y}, \mathrm{s}]|+) = P(\mathrm{y}|+) \cdot P(\mathrm{s}|+)$$

$$= \frac{0+1}{7+1\cdot 4} \cdot \frac{5+1}{7+1\cdot 2} = 0.0606$$

$$P(x = [\mathrm{y}, \mathrm{s}]|-) = P(\mathrm{y}|-) \cdot P(\mathrm{s}|-) = (\mathrm{s}|+)$$

$$= \frac{0+1}{5+1\cdot 4} \cdot \frac{3+1}{5+1\cdot 2} = 0.0635$$

$$P(+|x = [\mathrm{y}, \mathrm{s}]) = P(x = [y, s]|+)P(+)$$

$$= 0.0606 \cdot 0.58 = 0.0351$$

$$P(-|x = [\mathrm{y}, \mathrm{s}]) = P(x = [y, s]|-)P(-)$$

$$= 0.0635 \cdot 0.42 = 0.0267$$