



Naïve Bayes

Data Science: Jordan Boyd-Graber
University of Maryland

MARCH 1, 2018

Naïve Bayes

$$P(c|d) \propto$$

$$P(c) \prod_{1 \leq i \leq n_d} P(w_i|c) \quad (1)$$

$$\hat{P}(c_j) = \frac{N_c + 1}{N + |C|} \quad (2)$$

$$\hat{P}(w|c) = \frac{T_{cw} + 1}{\sum w' T_{cw'} + |V|} \quad (3)$$

Color	Type	Origin	Stolen
red	sports	domestic	Y
red	sports	domestic	N
red	sports	domestic	Y
yellow	sports	domestic	N
yellow	sports	imported	Y
yellow	suv	imported	N
yellow	suv	imported	Y
yellow	suv	domestic	N
red	suv	imported	N
red	sports	imported	Y

Naïve Bayes

$$P(c|d) \propto$$

$$P(c) \prod_{1 \leq i \leq n_d} P(w_i|c) \quad (1)$$

$$\hat{P}(c_j) = \frac{N_c + 1}{N + |C|} \quad (2)$$

$$\hat{P}(w|c) = \frac{T_{cw} + 1}{\sum_{w'} T_{cw'} + |V|} \quad (3)$$

Color	Type	Origin	Stolen
red	sports	domestic	Y
red	sports	domestic	N
red	sports	domestic	Y
yellow	sports	domestic	N
yellow	sports	imported	Y
yellow	suv	imported	N
yellow	suv	imported	Y
yellow	suv	domestic	N
red	suv	imported	N
red	sports	imported	Y

Estimate

$$\hat{P}(f = \text{red} | c = \text{stolen}), \hat{P}(f = \text{red} | c = \neg\text{stolen}), \hat{P}(t = \text{suv} | c = \text{stolen}), \\ \hat{P}(t = \text{suv} | c = \neg\text{stolen}), \hat{P}(o = \text{domestic} | c = \text{stolen}), \hat{P}(o = \text{domestic} | c = \neg\text{stolen})$$

Estimation

- $\hat{P}(f = \text{red} | c = \text{stolen})$

Estimation

- $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3+1}{5+2} = \frac{5}{7} \quad (4)$$

Estimation

- $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{5}{7} \quad (4)$$

- $\hat{P}(f = \text{red} | c = \neg\text{stolen})$

Estimation

- $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{5}{7} \quad (4)$$

- $\hat{P}(f = \text{red} | c = \neg\text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (5)$$

Estimation

- $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{5}{7} \quad (4)$$

- $\hat{P}(f = \text{red} | c = \neg\text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (5)$$

- $\hat{P}(t = \text{suV} | c = \text{stolen})$

Estimation

■ $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{5}{7} \quad (4)$$

■ $\hat{P}(f = \text{red} | c = \neg\text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (5)$$

■ $\hat{P}(t = \text{suv} | c = \text{stolen})$

$$\frac{1 + 1}{5 + 2} = \frac{2}{7} \quad (6)$$

Estimation

- $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{5}{7} \quad (4)$$

- $\hat{P}(f = \text{red} | c = \neg\text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (5)$$

- $\hat{P}(t = \text{suv} | c = \text{stolen})$

$$\frac{1 + 1}{5 + 2} = \frac{2}{7} \quad (6)$$

- $\hat{P}(t = \text{suv} | c = \neg\text{stolen})$

Estimation

■ $\hat{P}(f = \text{red} | c = \text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{5}{7} \quad (4)$$

■ $\hat{P}(f = \text{red} | c = \neg\text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (5)$$

■ $\hat{P}(t = \text{suv} | c = \text{stolen})$

$$\frac{1 + 1}{5 + 2} = \frac{2}{7} \quad (6)$$

■ $\hat{P}(t = \text{suv} | c = \neg\text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{4}{7} \quad (7)$$

Estimation

- $\hat{P}(o = \text{domestic} | c = \text{stolen})$

Estimation

- $\hat{P}(o = \text{domestic} | c = \text{stolen})$

$$\frac{2+1}{5+2} = \frac{3}{7} \quad (8)$$

Estimation

- $\hat{P}(o = \text{domestic} | c = \text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (8)$$

- $\hat{P}(o = \text{domestic} | c = \neg\text{stolen})$

Estimation

- $\hat{P}(o = \text{domestic} | c = \text{stolen})$

$$\frac{2 + 1}{5 + 2} = \frac{3}{7} \quad (8)$$

- $\hat{P}(o = \text{domestic} | c = \neg\text{stolen})$

$$\frac{3 + 1}{5 + 2} = \frac{4}{7} \quad (9)$$

Classification: red domestic SUV

$$P(c = \text{stolen} | d) \propto P(c) \prod_{1 \leq i \leq n_d} P(w_i | c) \quad (10)$$

$$= \frac{6}{12} \frac{5}{7} \frac{3}{7} \frac{2}{7} \quad (11)$$

$$= 0.0437 \quad (12)$$

Classification: red domestic SUV

$$P(c = \text{stolen} | d) \propto P(c) \prod_{1 \leq i \leq n_d} P(w_i | c) \quad (10)$$

$$= \frac{6}{12} \frac{5}{7} \frac{3}{7} \frac{2}{7} \quad (11)$$

$$= 0.0437 \quad (12)$$

$$P(c = \neg \text{stolen} | d) \propto P(c) \prod_{1 \leq i \leq n_d} P(w_i | c) \quad (13)$$

$$= \frac{6}{12} \frac{3}{7} \frac{4}{7} \frac{4}{7} \quad (14)$$

$$= 0.0700 \quad (15)$$