

Exercise 1.2

October 30, 2024

1 Estimate the probabilities

$$p(\text{yes}) = \frac{6}{10} \quad (1)$$

$$p(\text{red} \mid \text{yes}) = \frac{3}{6} = \frac{1}{2} \quad (2)$$

$$p(\text{grand tourer} \mid \text{yes}) = \frac{2}{6} \quad (3)$$

$$p(\text{domestic} \mid \text{yes}) = \frac{2}{6} \quad (4)$$

$$p(\text{no}) = \frac{4}{10} \quad (5)$$

$$p(\text{red} \mid \text{no}) = \frac{1}{4} \quad (6)$$

$$p(\text{grand tourer} \mid \text{no}) = \frac{2}{4} \quad (7)$$

$$p(\text{domestic} \mid \text{no}) = \frac{3}{4} \quad (8)$$

2 Predict the probability that a car with properties $x_1 = \text{red}$, $x_2 = \text{grand tourer}$, $x_3 = \text{domestic}$ will be stolen.

$$\begin{aligned} p(\text{yes}) \cdot p(\text{red} \mid \text{yes}) \cdot p(\text{grand tourer} \mid \text{yes}) \cdot p(\text{domestic} \mid \text{yes}) \\ = \frac{6}{10} \cdot \frac{1}{2} \cdot \frac{2}{6} \cdot \frac{2}{6} \\ = \frac{3}{40} \end{aligned}$$

3 What are the benefits, what are the downsides of using Naive Bayes?

The benefits are that it does not require a lot of data and is easy to implement, being ideal for smaller datasets. It is also quite fast.

The downside is that it is too dependent on data being of high quality/ Noisy data can result in incorrect results. It also assumes that all features are independent, which can limit its effectiveness where this is not the case.

4 The extra mile: Derive Equation 1 using Bayes' theorem, the chain rule of probabilities and the conditional independence assumption stated above.

Bayes' Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Chain Rule of Probability:

$$P(A \wedge B) = P(A) \cdot P(B)$$

Conditional Independence Assumption:

$$P(A \wedge B|C) = P(A|C) \cdot P(B|C) \quad \text{if A and B are conditionally independent given C}$$

Goal: Estimating conditional probability of y

$$P(y = k|x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n|y = k) \cdot P(y = k)}{P(x_1, x_2, \dots, x_n)}$$

Expanding the numerator using conditional independence:

$$= \frac{P(x_1|y = k) \cdot P(x_n|y = k)}{P(x_1, x_2, \dots, x_n)} \cdot P(y = k)$$

Using the product rule for conditional probabilities:

$$= \frac{1}{Z} P(y = k) \prod_{i=1}^n P(x_i|y = k)$$

with $Z = P(x) = P(x_1, x_2, \dots, x_n)$.