Naïve Bayes – Final Remarks

Dr. Muhammad Wasim

Simplification of Naïve Bayes Equation for Implementation

Naïve Assumption that the features are independent

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots, x_n \mid y)}{P(x_1, \ldots, x_n)} \Rightarrow P(y \mid x_1, \ldots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \ldots, x_n)}$$

Denominator is constant, so, lets ignore it

$$P(y \mid x_1, \ldots, x_n) \propto P(y)\prod_{i=1}^{n} P(x_i \mid y) \quad \Rightarrow \quad \hat{y} = argmax_y \quad P(y)\prod_{i=1}^{n} P(x_i|y)$$

Multiplication of many small numbers can cause underflow, so

$$\hat{y} = argmax_y \ \log(P(y)\prod_{i=1}^{n} P(x_i|y)\ ) \quad \Rightarrow \quad \hat{y} = argmax_y \ \log P(y) + \sum_{i=1}^{n} \log P(x_i|y)$$

Strengths and Weaknesses of Naïve Bayes

- **Strengths:**
  - Very fast to train and predict
  - Works quite well on sparse data (like text documents)
  - Can be used with very large datasets
- **Weaknesses:**
  - It assumes that all the features are independent. While it might sound great in theory, in real life, you'll hardly find a set of independent features.