5.07 Naïve Bayes Classifiers

Naïve Bayes

- A Machine Learning classification method that you can use to predict the likelihood that an event will occur given evidence that is present in your data

Naïve Bayes

If we know $P(B|A)$, Bayes theorem allows us to calculate the probability of $P(A|B)$ by relating the probability of $P(A|B)$ to $P(B|A)$.

$$\text{Bayes' Theorem: } P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Let A be that a message is spam.
- Let B represent the words used in the message.

$$\text{Bayes' Theorem: } P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\Rightarrow P(\text{message is spam}|\text{words in message}) = \frac{P(\text{words in message}|\text{message is spam})P(\text{message is spam})}{P(\text{words in message})}$$

We want to calculate the probability that a post is spam **given** the words that are in the message! Our model can learn this from the training data.

Naïve Bayes makes the assumption that all features are independent of one another (this is why it is called *naïve*).

Naïve Bayes Model Types

- Bernoulli - when we have 0/1 variables
- Multinomial - when our variables are positive integers
- Gaussian - when our features are Normally distributed

Naïve Bayes Use Cases

- Spam Detection
- Customer Classification
- Sentiment Classification

Naïve Bayes Assumptions

- Predictors are independent of each other

- Past conditions still hold true
  - When we make predictions from historical values, we will get correct results if the circumstances remain unchanged