Naïve Bayes

WILL STEARNS, SHARIF RAKHIMOV, PHIL CARBINO, DEREK PRESLAR

Naïve Bayes: Conditional Probability Basics

Conditional probability is used to calculate the probability of one event occurring given that another event has already occurred:

$$P(A|B)$$

Conditional probabilities are calculated in the following way:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad ; \quad P(A|B)P(B) = P(A \text{ and } B) = P(B|A)P(A)$$

From there, Bayes' Theorem can be derived:

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

Naïve Bayes: The Central Assumption

Suppose we have k different variables $x_1, x_2, \ldots, x_k$ from which we want to predict a target class, Y. Bayes' theorem can then be used to find the conditional probability of a particular y given that $x_1, x_2, \ldots, x_k$ are true:

$$P(y|x_1, x_2, \ldots x_k) = \frac{P(x_1,x_2,\ldots,x_k|y)P(y)}{P(x_1,x_2,\ldots,x_k)}$$

Naïve Bayes is based the assumption that all variables $x_1, x_2, \ldots, x_k$ are *independent*, meaning that no x value has any effect on the probability of another x value occurring, so we can split $P(A, B) = P(B)P(A)$. With this independence assumption the above equation becomes:

$$P(y|x_1, x_2, \ldots x_k) = \frac{P(x_1|y) * P(x_2|y) \cdots P(x_k|y) * P(y)}{P(x_1) * P(x_2) \cdots P(x_k)}$$

Naïve Bayes

A Naïve Bayes algorithm finds the most likely class y for the given data points $x_1, x_2, \ldots x_k$ (the y value which maximizes the conditional probability $P(y|x_1, x_2, \ldots x_k)$).

There are several different kinds of Naïve Bayes models, including:

- Gaussian Naïve Bayes
- Multinomial and Categorical Naïve Bayes
- Bernoulli Naïve Bayes
- Complement Naïve Bayes

Advantages:

- Simple to implement
- Fast and effective in predicting the class of datasets
- Doesn’t require a large test set to work well
- Can handle both discrete and continuous data
- Useful for text analysis and classification problems
- Can be used for multiple class prediction problems

Disadvantages:

- Relies upon the assumption that variables are independent
- Not ideal for data sets with a large number of variables
- Zero Frequency Problem

Data Processing Steps

- Convert categorical features to numeric (get dummies)
- Handle missing data. Drop or impute
- Identify outliers, remove or convert into numbers
- Separate dependent feature from independent features
- Gaussian- continuous data, need to be distributed normally
- Data smoothing- so that no probabilities are completely zero

Most Adjusted Hyperparameters

Gaussian Naïve Bayes

- var_smoothing

Bernoulli Naïve Bayes

- alpha
- fit_prior
- binarize

Multinomial and Categorical Naïve Bayes

- alpha
- fit_prior

Complement Naïve Bayes

- alpha
- fit_prior
- norm