Naive Bayes Classifier

Md. Abu Sadeed
ID - 160041017

July 13, 2020

Naive Bayes Classifier

Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from P(c), P(x), and $P(x|c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

- $P(c|x)$ is the posterior probability of class (target) given predictor (attribute).
- P(c) is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

Naive Bayes (NB) is 'naive' because it makes the assumption that features of a measurement are independent of each other. This is naive because it is (almost) never true.

Mathematical Derivation

Given a data point $x = x_1, ..., x_n$ of n features, naive Bayes predicts the class C_k for x according to the probability

$$P(C_k|x) = P(C_k|x_1, ..., x_n) \text{ for } k = 1, ..., K$$

Using Bayes' Theorem, this can be factored as

$$P(C_k|x) = P(x)p(x|C_k)P(C_k) = \frac{P(x_1, ..., x_n|C_k)P(C_k)}{P(x_1, ..., x_n)}$$

Using the chain rule, the factor $P(x_1, ..., x_n|C_k)$ in the numerator can be further decomposed as

$$P(x_1, ..., x_n|C_k) = P(x_1|x_2, ..., x_n, C_k)P(x_2|x_3, ..., x_n, C_k).....P(x_n|C_k)$$

At this point, the "naive" conditional independence assumption is put into play. Specifically, naive Bayes models assume that feature $x_i$is independent of feature x_j for $i \neq j$ given the class C_k Using the previous decomposition, this can be formulated as

$$P(x_1, ..., x_n|C_k) = \prod_{i=1}^{n} P(x_i|C_k)$$

Thus,

$$P(C_k|x_1, ..., x_n) = P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

Zero Probability

One of the disadvantage with Naive-Bayes is that if occurrences of a class label and a certain attribute value together is 0 then the frequency-based probability estimate will be zero, this is known as "Zero Probability Problem".

An approach to overcome this "Zero Probability Problem" in a Bayesian setting is to add one to the count for every attribute value-class combination when an attribute value does not occur with every class value.

Usage

- A classification technique based on Bayes' Theorem. Works best on two level features.
- Naive Bayes algorithms are mostly used in sentiment analysis, spam filtering, recommendation systems etc.
- This is faster KNN classification.

Code

The classifier model implementation can be found on this link: **GitHub Repository**.