Questions and Answers for Probabilistic Models

Bingzhang Zhu

December 2, 2021

1 Why Naive Bayes is called Naive?

We call it naive because its assumptions (it assumes that all of the features in the dataset are equally important and independent) are really optimistic and rarely true in most real-world applications:
we consider that these predictors are independent.
we consider that all the predictors have an equal effect on the outcome. (like the day being windy does not have more importance in deciding to play golf or not)

2 Can you choose a classifier based on the size of the training set?

If the availability of data is a constraint, i.e. if the training data is smaller or if the dataset has a fewer number of observations and a higher number of features, we can choose algorithms with high bias/low variance like Naïve Bayes and Linear SVM.

If the training data is sufficiently large and the number of observations is higher as compared to the number of features, one can go for low bias/high variance algorithms like K-Nearest Neighbors, Decision trees, Random forests, and kernel SVM.

3 How would you use Naive Bayes classifier for categorical features? What if some features are numerical?

We can use any kind of predictor in a Naive Bayes classifier. All we need is the conditional probability of a feature given the class, i.e., P(F | Class).

For the categorical features, we can estimate P(F | Class) using a distribution such as multinomial or Bernoulli.

For the numerical features, we can estimate P(F | Class) using a distribution such as Normal or Gaussian.

For a numerical feature that follows a different specific distribution, for example, an exponential, then that specific distribution may be used instead.

For a numerical or categorical without a well-defined distribution, then a kernel density estimator can be used to estimate the probability distribution instead.

Since Naive Bayes assumes the conditional independence of the features, it can use different types of features together. We can calculate each feature's conditional probability and multiply them to get the final prediction.