CSE 575 - Project 1 Part 1 - Naive Bayes Classifier - Report

ABHISHEK PATIL
Student ID: 1216126522

INTRODUCTION

The Naive Bayes Classification is a popular technique in machine learning and is based on the famous Bayes theorem. The theorem is given by the following formula for two features/attributes:

$$p(y|x_1,\ x_2) = \frac{p(x_1, x_2|y)\cdot P(y)}{\sum_{y}[p(x_1, x_2|y)\cdot P(y)]}$$

Eq.1. Bayes Rule probability of event y given x_1 and x_2

We use the classifier to classify digit images as either 0 or 1. We are given training and testing datasets for both 0 and 1 images in the form of 2D arrays of pixel values. As per the formula in Eq.1, event y is the digit value of 1 or 0. x_1 and x_2 are two features: mean (feature 1) and standard deviation (feature 2) of pixel brightness values.

RESULTS

8 Parameters are:

1. Mean of feature1 for digit0 :	44.17427270408163
2. Variance of feature1 for digit0 :	114.4469469686328
3. Mean of feature2 for digit0 :	87.40062256913791
4. Variance of feature2 for digit0 :	101.60057622593692
5. Mean of feature1 for digit1 :	19.44723775510204
6. Variance of feature1 for digit1 :	31.585043570284256
7. Mean of feature2 for digit1 :	61.47723875716028
8. Variance of feature2 for digit1 :	82.68073379891041

2 Accuracy values are:

1. Accuracy for digit0testset :	0.9173469387755102
2. Accuracy for digit1testset:	0.9233480176211454

OBSERVATION AND UNDERSTANDING

The Naive Bayes classification is very easy to implement but can be tough to get a grasp on. Once the idea behind posterior and prior probabilities is clear, implementing the classifier is not tough. However, the experience provided in this rudimentary project is crucial to begin implementing complex problems with the same method

The advantage of this method is that it allows us to express the probability of an event given some attributes in the form of posterior probabilities: probabilities of the different features given the event. Getting posterior probabilities is as simple as finding the relative frequencies in the table or by assuming a particular distribution and using its probability distribution function.

The problem statement and algorithm make the assumption that the dataset comes from a normal distribution. Thus, mean and variance of the dataset is calculated and the normal distribution formula is used to make the estimates on the posterior probabilities. Such assumptions on the dataset depends on the problem scenario and having some domain knowledge and past experience with similar data is a huge advantage. The expression for getting the probability given a test data point x, mean μ and variance σ^2 is shown below in eq.2.

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} * e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Eq.2. Normal probability distribution for a given mean μ and variance σ^2

Another simplifying assumption made in this dataset is that the features are conditionally independent from each other given the event y. Such an assumption allows us to reduce the number of probability values required to make a prediction. As a consequence, we only have to get the probability of a specific attribute given the class label while not caring about how other attributes affect it. We also may not have such information available in some cases. In this example however, having only 2 dimensions does not show the true significance of the assumption.

It is important to appreciate how an accuracy of above 90% is achieved for both the 0 and 1 digit test data set even with the simplifying assumption. This kind of classification gets estimates on the likelihood of a certain event y based on probabilities coming from the dataset. If the dataset has a lot of attributes, the assumption makes it computationally possible to make predictions without losing too much.

To conclude, Naive Bayes classification is a very intuitive way of tackling the problem of classification. Whenever we are faced with a decision to make, we look at some history of data and try to predict outcomes by comparing new (test data) circumstances with previous (train data) experiences. The disadvantage is that if we face unseen/untrained circumstances (test data) the decision making is not helpful, which is also the case of Nayes Bayes where the probability value is just 0 (never happened before).