Naïve Bayes Classifier Project

1. Introduction

In this project, we are going to use MNIST dataset for training and testing and use Naïve Bayes Classifier to predict. We are using only digit 0 and digit 1. We are going to learn about the Bayes Theorem and how it is used in our case. And then we are going to talk about features we need and how to implement them to Naïve Bayes Theorem. And then we are going to talk about Bayesian Inference and Gaussian Naïve Bayes for posterior probability and conditional probability calculation. At last, we will see the result of using them to calculate the accuracy of our formula.

2. Bayes Theorem to Classify

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

In our case, $P(A|B)$ becomes $P(Y = 0|X)$ and $P(Y = 1|X)$ for digit 0 and digit 1. We have image files for digit 0 and digit 1. There are 5000 images of training set for digit 0 and 5000 images of training set for digit 1. We have 980 images of test set for digit 0 and 1135 images of test set for digit 1. We calculate mean and variance of features. The two features are the mean and standard deviation of the brightness of each image file of digit 0 training set and digit 1 training set. We will have two features for each training set. We assume all features are mutually independent and we use Naïve Bayes Theorem to calculate the posterior probability.

$$X = (x_1, x_{2,}\ x_3, \ldots, x_n)$$

X is the set of the features and we will have only $x_1\ and\ x_2$, since we have only two features. And then we plug the features in the Naïve Bayes formula below.

$$P(Y \mid X) = \frac{P(x_1|\ Y)\ \cdot\ P\ (x_2|Y) \cdot \ldots \cdot\ P(x_n|Y)\ \cdot P(Y)}{P(X)}$$

We will need to know about the likelihood of the Gaussian Naïve Bayes and its formula to calculate the conditional probabilities of Naïve Bayes Theorem.

3. Bayesian inference

$$P(Y|X) \propto P(Y) \cdot P(X|Y)$$

$P(Y|X) \cdot P(Y)$ is proportional to $P(Y|X)$. We can use the conditional probability and the prior probability to calculate the posterior probability, $P(Y|X)$. This makes the calculation simpler.

4. Gaussian Naïve Bayes

$$p(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

We can calculate the conditional probabilities for the posterior probability calculation by using the likelihood formula of Gaussian Naïve Bayes. We use the Gaussian Naïve Bayes since the dataset is continuous data.

We calculate the mean and variance of each feature we calculated. the mean and variance of the mean of digit 0 are the mean of feature 1 and the variance of feature 1. We assign them to variables like *mean_fit1_train0* and *var_ft1_train0*. The mean and variance of the standard deviation of digit 0 are the mean and the variance of feature 2 and assign them to variables like *var_ft1_train0* and *var_ft2_train0*. We do the same process for the digit 1 dataset for training.

Now we have variables like *mean_ft1_train0, var_ft1_train0, mean_ft2_train0, var_ft2_train0, mean_ft1_train1, var_ft1_train1, mean_ft2_train1, var_ft2_train1 for digit 0 and digit 1 dataset for training*. Now we are ready to calculate the conditional probabilities that are required by the posterior probability calculation.

5. Posterior Probability Calculation and Prediction

We can use the formula from Gaussian Naïve Bayes to calculate the conditional probabilities of each feature of digit 0 and digit 1. Probability of priors of digit 0 and digit 1 are same since the number of samples of digit 0 and digit 1 are same. If we are calculating the probability of $P(Y_0|X)$, we multiply $P(x_1|Y_0) \cdot P(x_2|Y_0) \cdot P(Y_0)$. We can use log for the multiplication because the probabilities are small and may cause the overflow problem and when you use log the multiplication becomes summation, which is easier. If we are predicting using digit 0, we are going to predict digit 0 using the digit 0 test dataset and the features from training set from digit 0 and digit 1. We will have two probabilities from predicting each image of digit 0 using the features from digit 0 and digit 1 training set. We can compare the probabilities and higher probability means that it is what we predict. If the features from digit 0 training set give higher probability than it of digit 1, then we are predicting the image is digit 0. We can use argmax or simply compare the two probabilities and calculate the accuracy.

6. Conclusion

We calculated the mean and variance of two features from each digit. We calculated the posterior probabilities using Naïve Bayes and Gaussian Naïve Bayes Theorems. The values of the eight parameters that contain the mean and variance of the two features of each digit are in sequence of [Mean_of_feature1_for_digit0, Variance_of_feature1_for_digit0, Mean_of_feature2_for_digit0, Variance_of_feature2_for_digit0, Mean_of_feature1_for_digit1, Variance_of_feature1_for_digit1, Mean_of_feature2_for_digit1, Variance_of_feature2_for_digit1] and are following:
[44.291450510204086, 116.08974396799537, 87.496082609966706, 102.06474666443756, 19.404539540816327, 31.797745546320222, 61.401555198985854, 83.222985900306114]

The accuracy calculation for digit 0 and digit 1 with test0 and test1 are following:
Accuracy for digit0 with test0: 0.9173469387755102
Accuracy for digit1 with test1: 0.9233480176211454