NAÏVE BAYES CLASSIFIER

Machine Learning Assignment 3

Submitted by,

Amala Deshmukh –	**2014A7PS541H**
Bhargav Kanuparthi-	**2014A7PS527H**
Karthik Menon –	**2013B3A7487H**
Keyur Jain -	**2015A7PS056H**

Naive Bayes Classifier

Methodology:

The task of classifying whether a given image is a face or not was done using Naive Bayes classifier, where each input feature was one corresponding pixel in the image.

The images pixel value is either '#', corresponding to an edge being found at that location, or ' ', corresponding to a non-edge pixel. The dimensions of each input image is **70x60**.

Now the following probabilities were calculated.

- **P(face)** – probability that a given image corresponds to a face int the dataset.
- **P(not face)** – probability that a given image does not correspond to a face.
- **P(pixel='#'|face)** – the probability a given pixel contains a # given that the image was of a face. The same probability was computed for images which are not faces i.e. **P(pixel = '#'|not face)**.
- Similarly **P(pixel = ' '|face)** and **P(pixel = ' '|not face)** were also computed.

After the probabilities were computed , we calculate if the **posterior probability** is more for it being a face or non face and give the output. But there was a problem while calculation this as the probabilities were converging to 0. To handle this at each iteration the probability was multiplied with a scaling factor of 1.5 so that they don't converge to 0.

Also a smoothing factor of 1 was used to prevent 0 probability from happening.

Snapshot of the training set accuracy , testing set accuracy and Confusion Matrix.

```
The training set accuracy is 95.121951
The testing set accuracy is 90.666667


Classification Matrix

68.000000 68.000000
9.000000 5.000000


Confusion matrix
88.311688 93.150685
11.688312 6.849315
```

Interesting examples of false positives(Classified as face , when actually not a face)

- **17th** input in the testing data

- **57th** input in the testing data

Interesting examples of false negatives(Classified as not a face , when actually it is a face)

- **27th** input in the testing data

- **30th** input in the testing data