Machine Learning
BITS F-464

Assignment #3

* * * * * * * *

Classification of data-set using Naïve-Bayes Classifier

(to infer the presence of a face)

* * * * * * * *

Submitted to:
Dr. N.L Bhanumurthy
(Dept. of Computer Science/ Information System)

* * * * * * * *

By
Bhavya Mishra : 2013B3A7657H
Mohit Jain : 2013B3A7675H
Manaswini Pandey : 2013B3A7719H
Ridam Jain : 2013B5A7841H

Training Of the Classifier:

- We use a **supervised learning technique**.
- **Encoding the entire input** in a matrix of in binary format **1s ('#') and 0s(' ')** : into a **train_input [train_size][R][C],** along with the **train_target[ ]** to store the target values. (1 to contain face and 0 for not containing face).
- Every element in this matrix is perceived as a predictor in itself.
- We use the arrays **P1yes, P1no, P0yes, and P0no** to act as the likelihood tables
- To solve the **zero - frequency problem**, we use the technique **'Laplace smoothing'** with the **pseudocount** value as **1**, which comes straight from the **laplace's Rule of Succession.**
- Each probability is calculated as:
  - **P1yes[i][j]** = (Number of faces with [i][j]th predictor is '#' + LP) / No. of faces
  - **P1no[i][j]** = (Number of faces with [i][j]th predictor is '#' + LP) / No. of faces
  - **P0yes[i][j]** = (Number of non-faces with [i][j]th predictor is ' ' + LP) / No. of non-faces
  - **P0no[i][j]** = (Number of non-faces with [i][j]th predictor is ' ' + LP) / No. of non-faces
    - where **LP = Pseudocount for Laplace Smoothing = 1**

Confusion Matrix

Accuracy obtained = **89.3333%**

True positives = 68	False positives = 11
False negatives = 5,	True negatives = 66

Examples Of Incorrect Classification

1. **False Positives**: false_positive_images.txt
2. **False Negatives**: false_negative_images.txt

These examples can be found in the attached files named as mentioned above.

Code

The code can be found in the attached file named as mentioned below:

Naive Bayes Classifier.cpp