Mini Project Report

Naïve Bayes Classification

Project Goal:

The main objective of this project is to design a finger print spoofing detector using Naïve Bayes classifier.

Data Set Used:

- featureMat_liv_train_bioLBP
- featureMat_Latex_train_bioLBP
- featureMat_liv_test_bioLBP
- featureMat_Latex_test_bioLBP
- featureMat_Gelatine_test_bioLBP

Basis of Analysis:

The model designed using the training set is evaluated using the test set based on the loss and re-substitution errors computed.

Findings:

Upon fitting a model on the **training data**, the **loss (classification error) is 0.1** and **re-substitution error is 0.1**. This means the model is a good fit and the training data is well spread across the feature vector and is well distributed.

The **live and latex test data** is used to validate the model and the resulting **loss is 0.2008** and the **re-substitution error is 0.1**. The classification error (loss) has increased but the change is minor, along with no change to the re-substitution error resulting in the model being a good fit. This leads to the classification of the data into the correct class. This results in our model being robust with high level of accuracy of the classifier.

Next, the **gelatin test data is also included**, and the **revised loss is 0.2067** and **re-substitution error is 0.1**. Adding more test data did not impact the accuracy of the classifier making the model robust and efficient.

Now, the **prior probabilities are updated to [0.6 0.4]** and the **loss and re-substitution errors** are computed on live and latex test data. The resulting values are **0.3606 and 0.1036 respectively**. We observe the loss (classification error) increases with a change in the priors of the classes. However, it does not impact our re-substitution error significantly. With the loss, though increased, to be still in a lower range.

Thus, with low loss and re-substitution error rates, we may **conclude** the **detector designed** may be said to be **robust and dependable in classifying the data accurately**.