CS 697AB Machine Learning

Graduate Assignment-2 R527Y884

Problem:

Figure print spoof data detection:

The problem we must solve in this assignment is we have four data sets given segregated into two modules 'Training and Testing' Data sets.

The Training Data sets again contain two separate data sets in them which are "live" and "spoof" data sets.

The Testing Data sets again contain two separate data sets in them which are "live" and "spoof" data sets.

We must build an SVM model (Support Vector Machine) to classify the live and spoof data and find whether our model is perfectly built based on our Scores and precisions values.

Process:

Step 1:

1. Importing the libraries from the specified packages.
2. Reading the data files from the machine (respective computers location).
3. Every image in the data has different dimensions. To fit an SVM model every image should be stabilized(ravel).
4. Every Machine Learning model can be predicted on numerical data only, so we get the array values of each image in our data.

Step 2:(Training the model)

5. Since we are having two data sets of training data (live and spoof). We concatenate them into a single list named "train_data".
6. Since we are performing a machine learning algorithm on the training data we need to classify the live and spoof data from the "train_data" based on the classes.
7. Let's define a classes for both (live = 1, spoof = 0). What this does is it divides the images of live to 1 and spoof to 0 in our target data which we create for building the model.

Step 3: SVM

Support Vector Machine is a Supervised Machine Learning algorithm that is used for classification and regression. It is used in the classification of problems. For

instance, assume our problem has both live and spoof data values. SVM classifies both separately and these are separated by a hyperplane.

The maximum distance of the data values of both live and spoof to hyperplane are called “support vectors”.

The hyperplane is defined as “C” in our problem which is used in classifying our data into live and spoof.

Using the training data, we got the

Step 4:(Testing the model on test data)

As our data is already trained, we now need to pass our testing data into the to check the accuracy of our model on the testing data. We do the same process for merging the live and spoof data into one list as we did earlier in our training data. And now we create a target_test(live=1, spoof=0).

After running our testing data into the model our score are :

- Model score for testing data = 89.0
- Recall score is 0.99
- Precision score is 0.78