

Submitted by

Debonil Ghosh (M21AIE225)

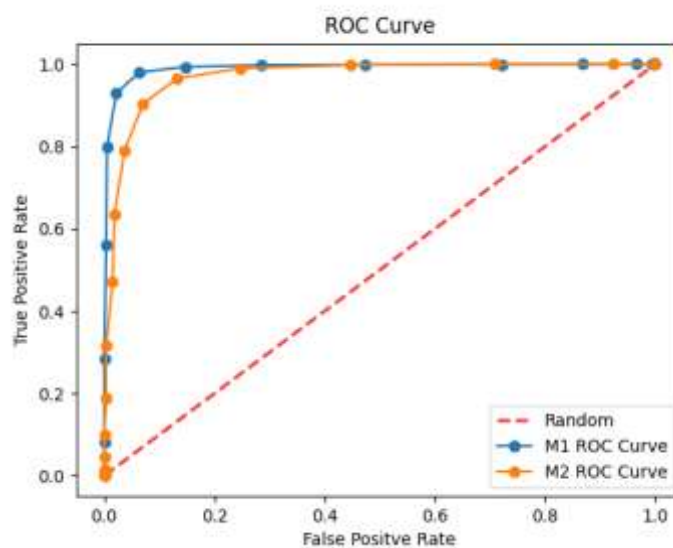
**Question 1.**

There are two models, M1 and M2, used to predict the scores for some input data. Suppose M1 predicts the score for input data as score1.npy and M2 predicts the score for the same data as score2.npy. Actual labels for a given score is label.npy (use np.load to load .npy files)

1. Plot ROC curve (from scratch) for both the models in a single plot. (10 marks)
2. Explain which model performs better on this data and why? (5 marks)
3. Compute AUC for both the ROC curves. (5 marks)
4. Calculate true positive rate for both models when false acceptance rate is 10% (5 marks)
5. Draw your analysis on (3) and (4) (5 marks)

Note: Scores here represent the distance between two samples using two different models. 0 in the label represents similar samples and 1 represents different samples.

**Solution:**



1.

**2. Model M1 performs better on this data.**

The curve that reaches closer to the top left corner of the graph denotes the better performing model. By visually comparing ROC graphs of given two models, it is found that ROC graph of M1 is slightly in top-left position with ROC curve of M2. It confirms that M1 will always give better (lesser False Positive points compare to True positive points) or same results than M2.

It is found that AUC of M1 is greater than AUC of M2 and that also supports the above statement.

**3.**

Area Under the curve (AUC) for M1: **0.9894237454500838**

Area Under the curve (AUC) for M2: **0.9681889933184838**

**4. When false acceptance rate is 10%,**

True positive rate for M1: **0.9858447507259417**

True positive rate for M2: **0.9334093356781591**

**5. Model M1 and M2 can achieve**

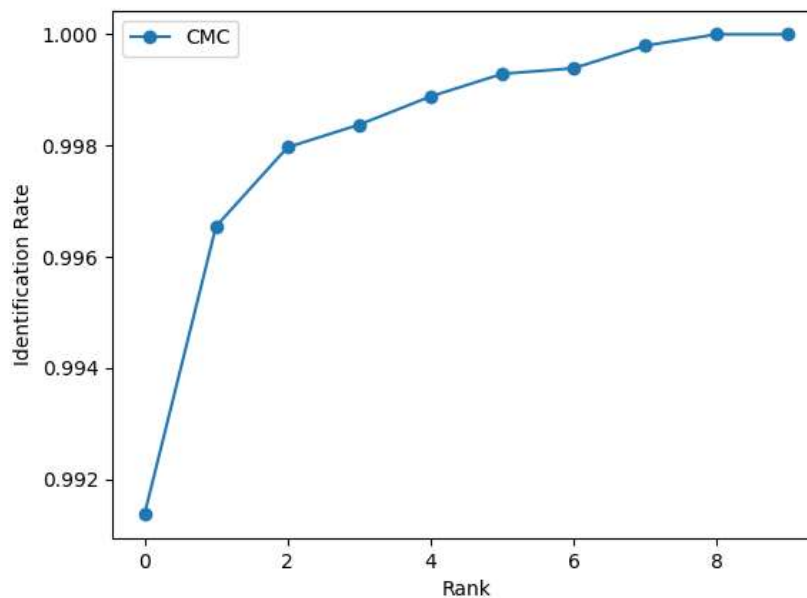
## Question 2.

Dataset link: [Link](#)

Consider a fingerprint recognition dataset, having 600 images in the gallery and 9854 images in the probe. A model is used to classify probe images into 600 classes. The probabilities predicted by the model for all 600 gallery images are given in `score.npy`. The correct labels are given in `label.npy`.

1. Plot CMC curve up to rank 10. (10 marks)
2. Comment on the results

**Solution:**



- 1.
- 2.

### Question 3.

You are requested to solve the fruit classification problem based on the features in the given dataset using decision trees. Load this dataset for your decision tree classification problem. The dataset has 3 features and one target variable. The target variable takes either Papaya (0) or Banana (1). The features are "Size" in cm, "Weight" in kg, and "SkinColor" (100-green, 200-yellow, and 300 orange).

- Load (Train-Test Split) and prepare required packages and shuffle the dataset. (2 marks)
- Build and Train a DecisionTree classifier. (5 marks)
- Don't stick to a single configuration for your model. Try different hyperparameters. (At least 5) (3 marks)
- Test the model for each configuration (5 marks)
- Visualize the tree, evaluate it based on the metrics given in previous questions. (3 marks)
- Report the confusion matrix for your best model (don't use inbuilt function) (2 marks)

If the hyperlink doesn't work copy-paste the URL below -

<https://drive.google.com/file/d/1O-Txgca54gFn0cTszrYq3n7OIKnz5o2m/view?usp=sharing>

**Solution:**