PATTERN RECOGNITION

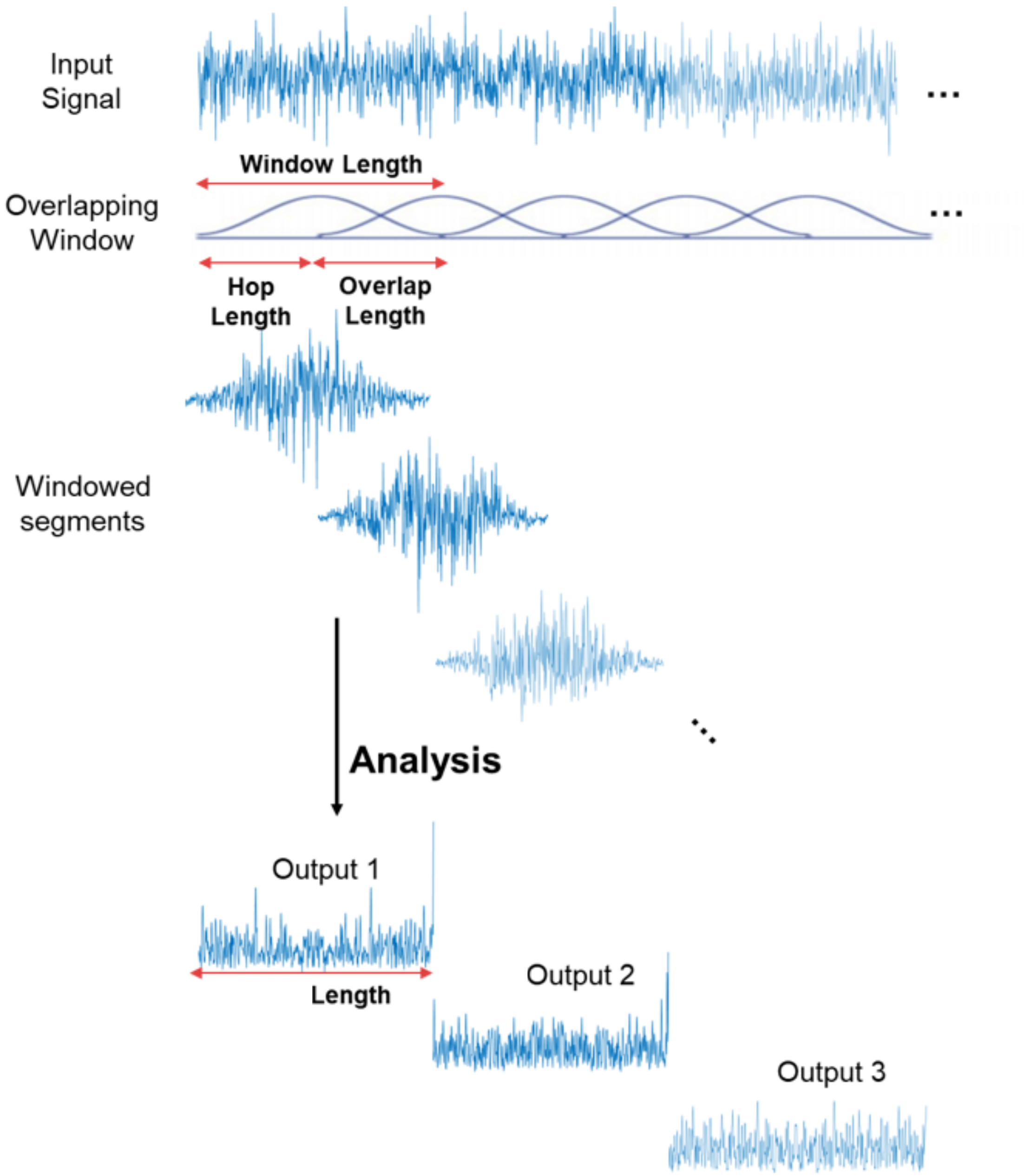
Assignment I
28 September 2022

SPEECH ACTIVITY DETECTION

Given an audio signal, the task is to classify the given frame of the signal is either speech or non-speech. Two features were given,

- Short Time Energy (STE)
- MEL Filterbank Energy

These are extacted from the audio file by windowing technique as shown below.



This figure describes the feature extraction process of an audio. For example in STE each energy is calulated for each outputs.

With the given feature, the task is to identify which of these features are better at correctly detecting speech.

MAXIMUM LIKELIHOOD FOR NORMAL DISTRIBUTION

A simple unimodal Gaussian to estimate the distribution of the features. The task is to find the mean and variance for estimation. The liklihood function is given by,

$$L(\mu, \sigma \mid x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Calculation of mean μ:

$\dfrac{\partial L(\mu, \sigma \mid x)}{\partial \mu} = 0$ by treating σ as constant.

For diferential simplicity, taking $ln(\,.\,)$ on both sides & rearranging we get,

$$ln(L(\mu, \sigma \mid x)) = ln(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x_1-\mu}{\sigma})^2} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x_2-\mu}{\sigma})^2}) + \ldots\ldots)$$

$$ln(L(\mu, \sigma \mid x)) = -\frac{n}{2} ln(2\pi) - n\,ln(\sigma) - \frac{1}{2}(\frac{x_1-\mu}{\sigma})^2 - \ldots\ldots$$

$$\frac{\partial ln[L(\mu, \sigma \mid x)]}{\partial \mu} = 0 - 0 + (\frac{x_1-\mu}{\sigma^2}) + \ldots\ldots = 0$$

$$\frac{1}{\sigma^2}(x_1 + x_2 + \ldots . x_n) - n\mu = 0$$

$$\mu = \frac{x_1 + x_2 + \ldots . x_n}{n}$$

which is the sample mean of the data points. So we need to calculate the mean of training examples.

Calculation of Variance σ^2:

$\dfrac{\partial L(\mu, \sigma | x)}{\partial \sigma} = 0$ by treating μ as constant.

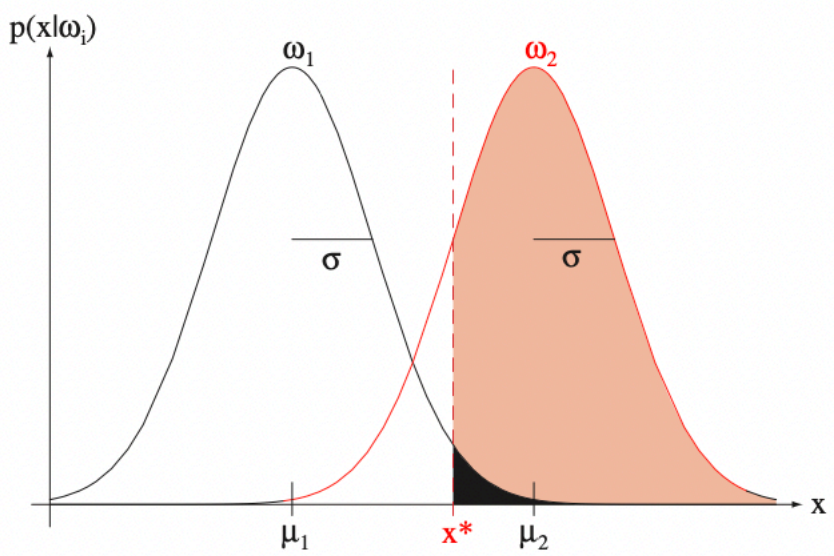
Similarly,

$$\frac{\partial ln[L(\mu, \sigma | x)]}{\partial \sigma} = 0 - \frac{n}{\sigma} + \frac{(x_1 - \mu)^2}{\sigma^3} + \ldots$$

$$-\frac{n}{\sigma} + \frac{(x_1 - \mu)^2}{\sigma^3} + \ldots = 0$$

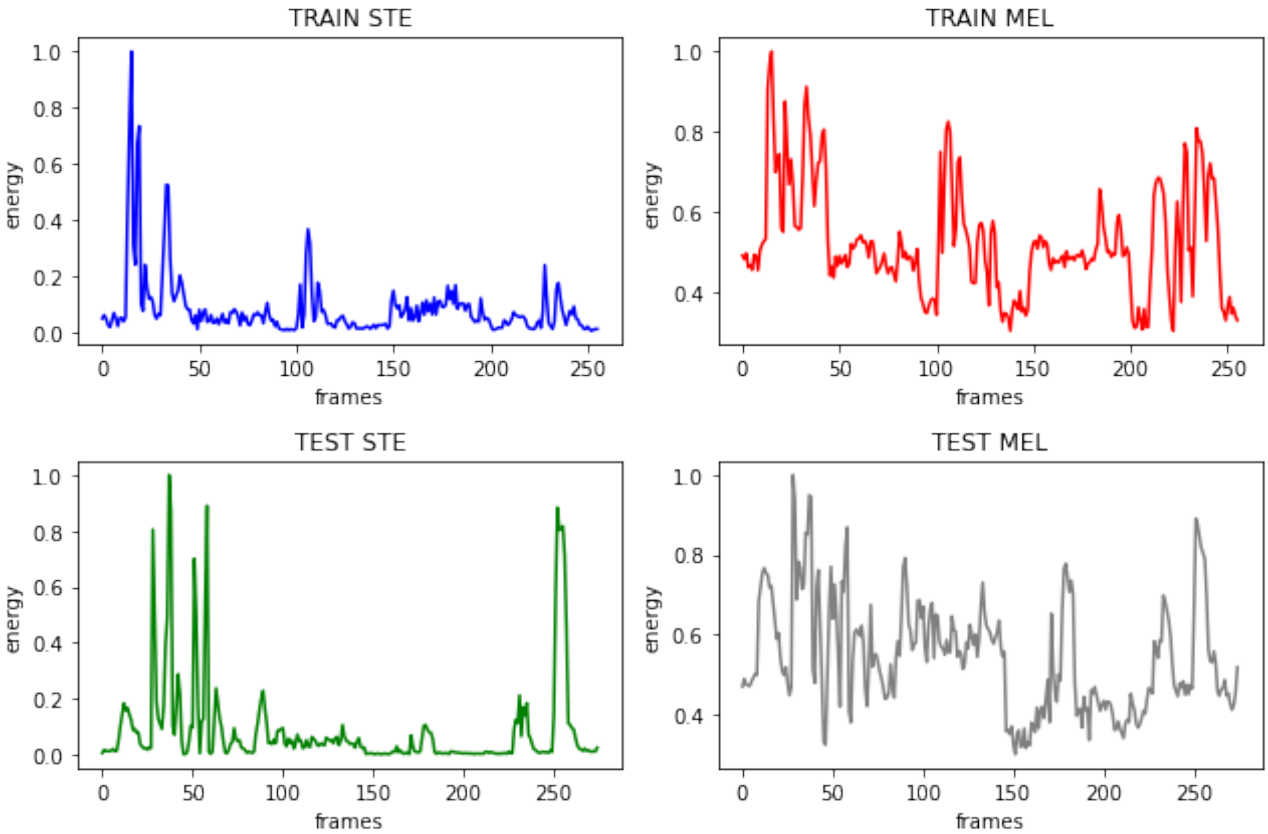
$$\sigma^2 = \frac{(x - \mu)^2 + \ldots (x_n - \mu)^2}{n}$$

which is the sample variance of the data points. So we need to calculate the variance of training examples.

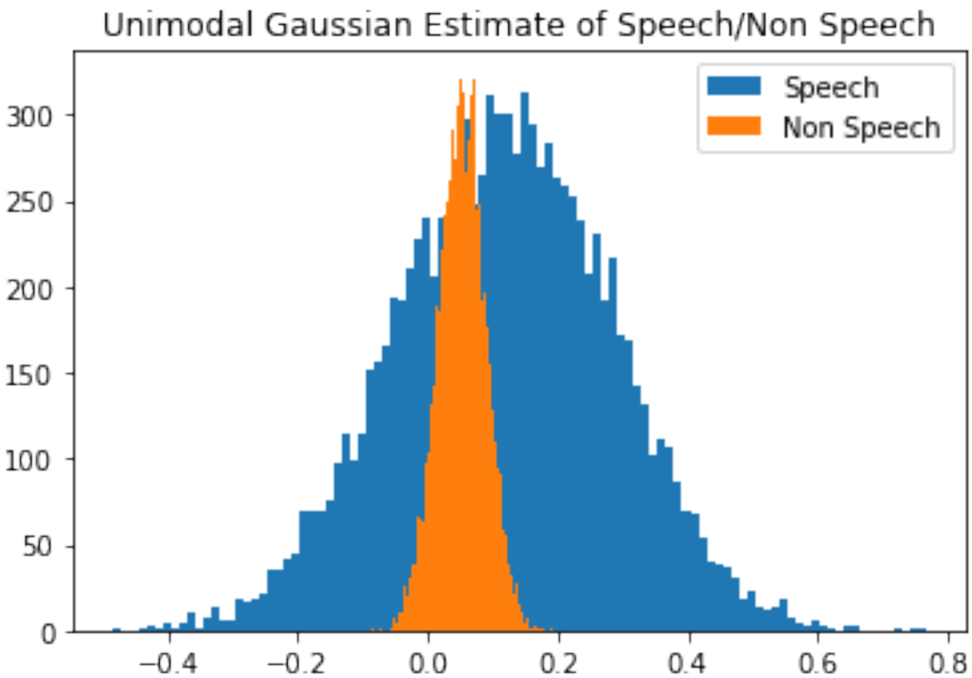


After estimating the mean and variance of the speech and non-speech dataset, we can fix the distribution for futher calculations for classifications. The plots theoretically will appear as shown. It is shown with different mean and same standard deviation.

STE & MEL FEATURES FOR TRAINING AND TEST DATASET



STE FEATURES



POSTERIOR FOR CLASSIFICATION

Here the classification of speech or non-speech is based on the posteriori which is given by,

$$p(s \mid x_i) = \frac{p(x_i \mid s)p(s)}{p(x_i)}$$

Where $p(\,.\,)$ is the probability, s is the speech samples, x_i is the test samples, $p(x_i \mid s)$ is the likelihood, $p(s)$ is prior of speech and $p(x_i)$ is the evidence. Since we are going to compare each posterior, $p(x_i)$ is not needed.

Finally by the Bayes theorem,

$$p(s \mid x_i) = \frac{p(x_i \mid s)p(s)}{p(x_i \mid s)p(s) + p(x_i \mid ns)p(ns)}$$

RECEIVER OPERATING CHARACTERISTICS (ROC)

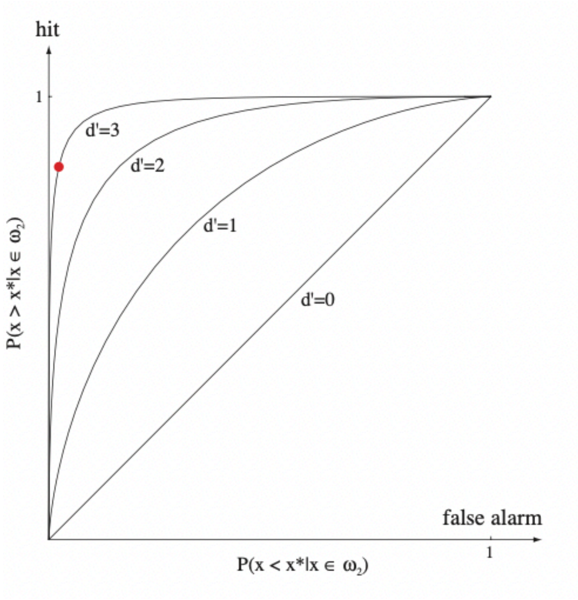
The ROC curve tells us the which feature gives us the better model by calculating the TPR (True Positive Rate) and FPR (False Positive Rate). Assume some threshold x^*,

$P(x > x^* | x \in \omega_2)$: a *hit* — the probability that the internal signal is above x^* given that the external signal is present

$P(x > x^* | x \in \omega_1)$: a *false alarm* — the probability that the internal signal is above x^* despite there being no external signal is present

$P(x < x^* | x \in \omega_2)$: a *miss* — the probability that the internal signal is below x^* given that the external signal is present

$P(x < x^* | x \in \omega_1)$: a *correct rejection* — the probability that the internal signal is below x^* given that the external signal is not present.



Iternating the threshold x^* for some 1000 values between the range 0 to 1 and calculating TPR and FPR for each values and plotted as graph called Reciever Operating Characteristics ROC curve.

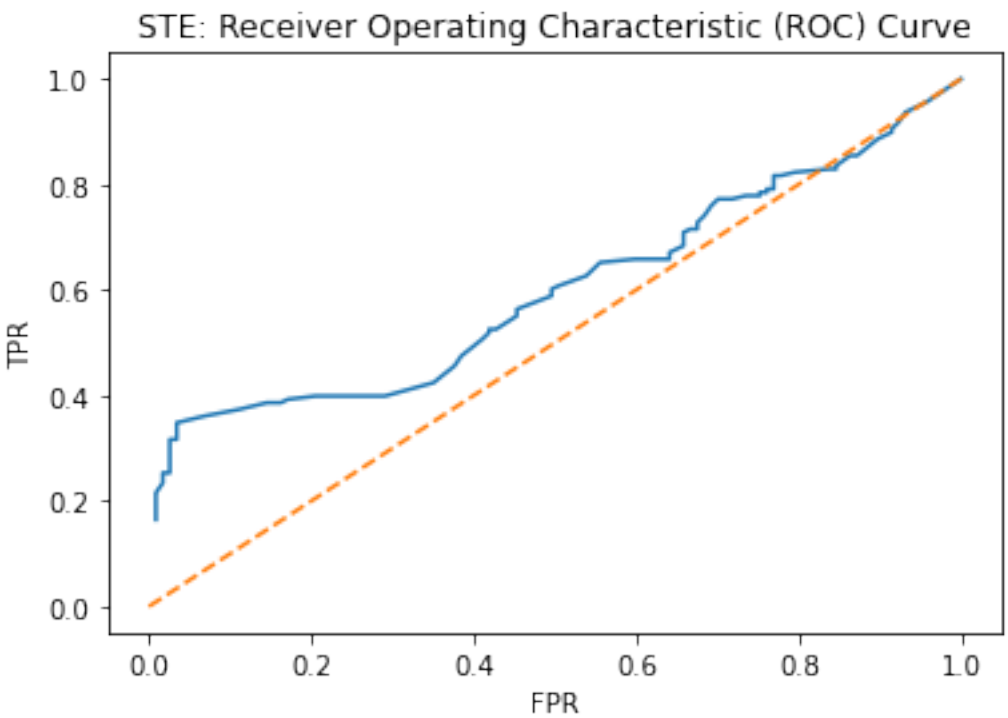
The ROC curve tells us at which threshold x^* and for what feature the model works well.

FORMULAS USED:

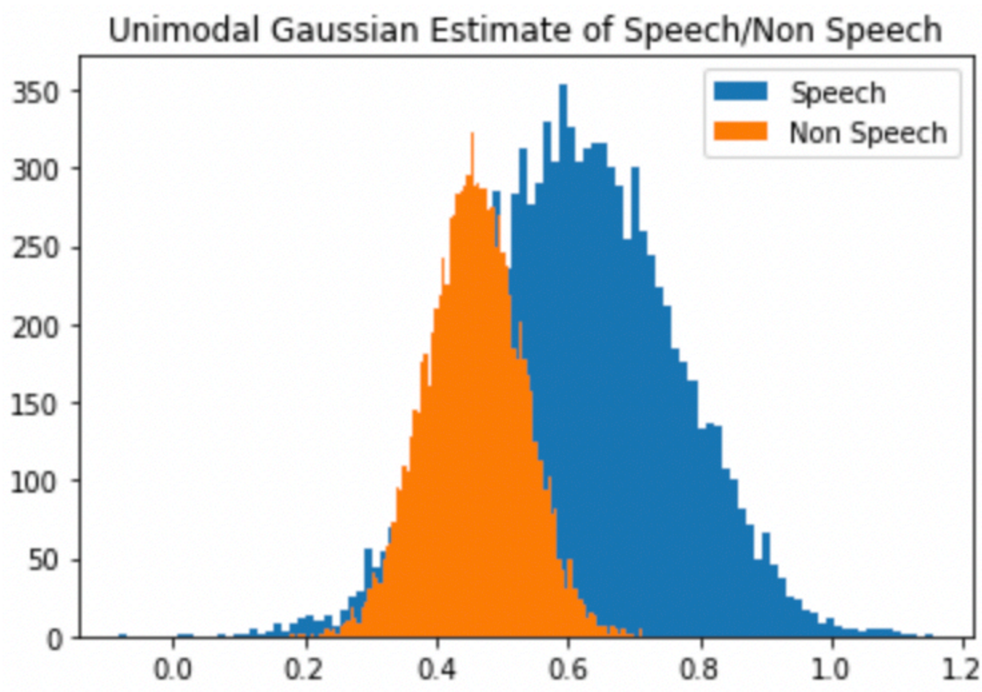
$$TPR \quad (or) \quad sensitivity = \frac{TP}{TP + FN}$$

$$FPR \quad (or) \quad 1 - selectivity = \frac{TN}{TN + FP}$$

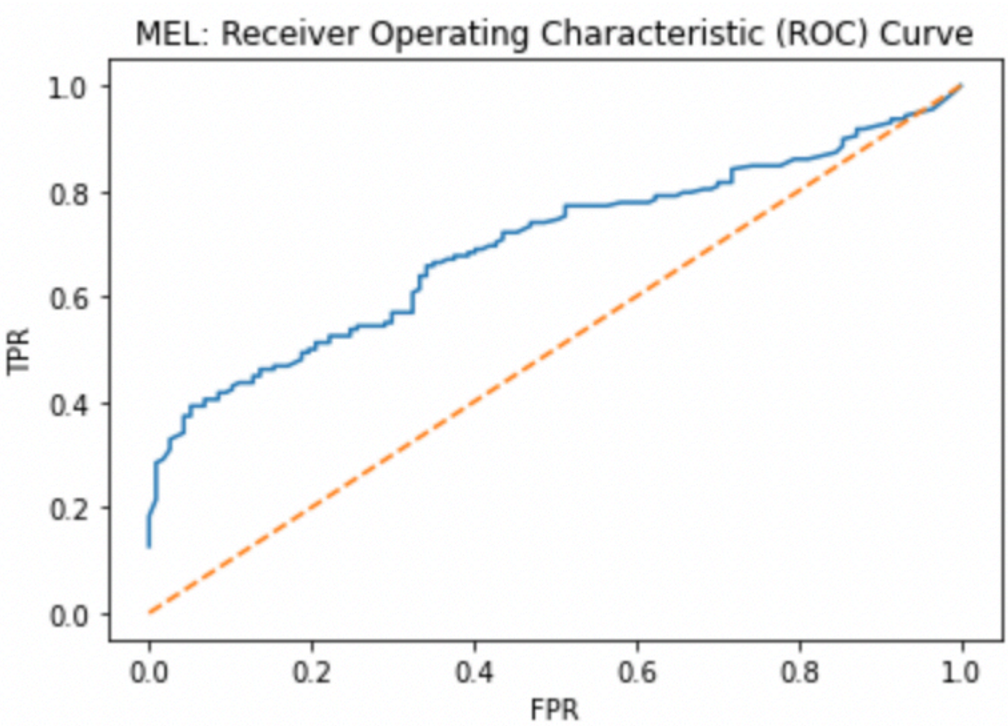
RESULTS (STE FEATURE)



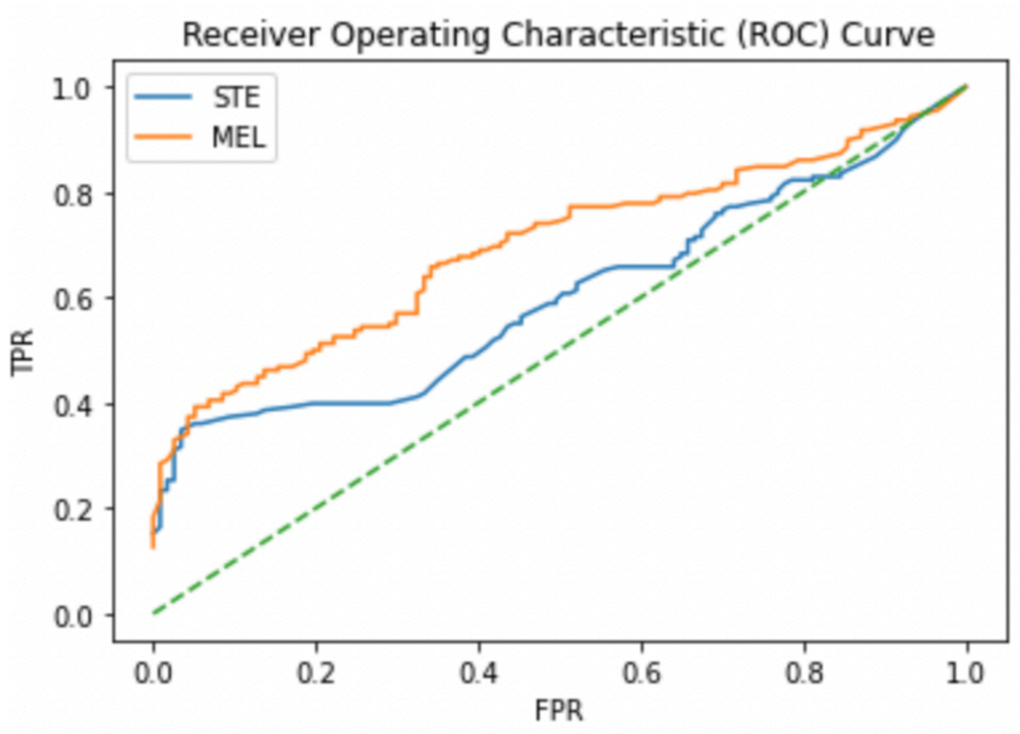
MEL FEATURES



RESULTS (MEL FEATURE)



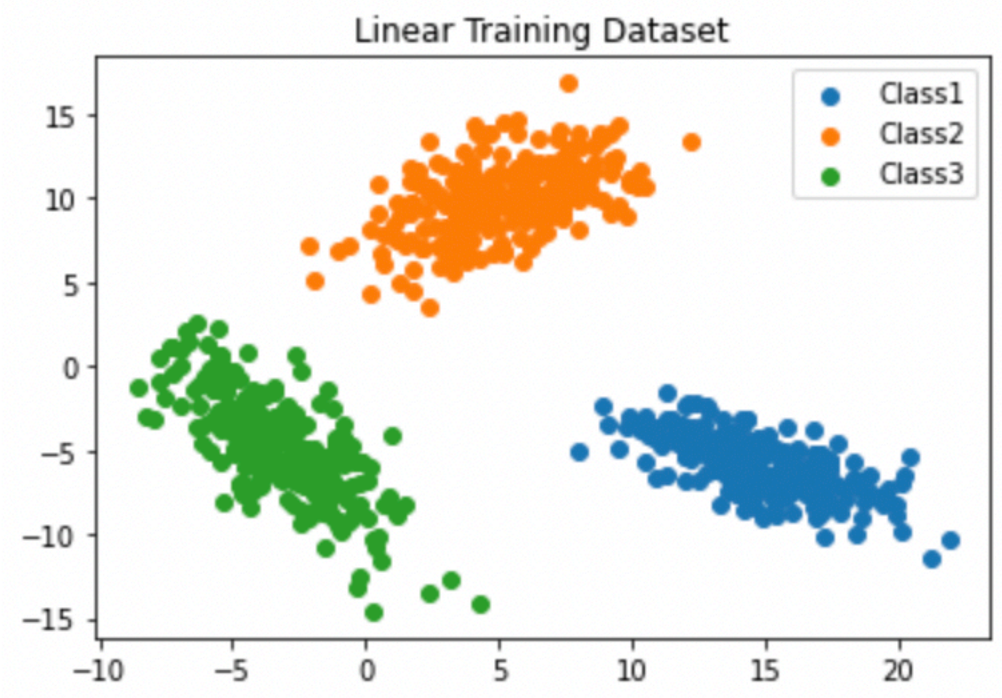
CONCLUSION

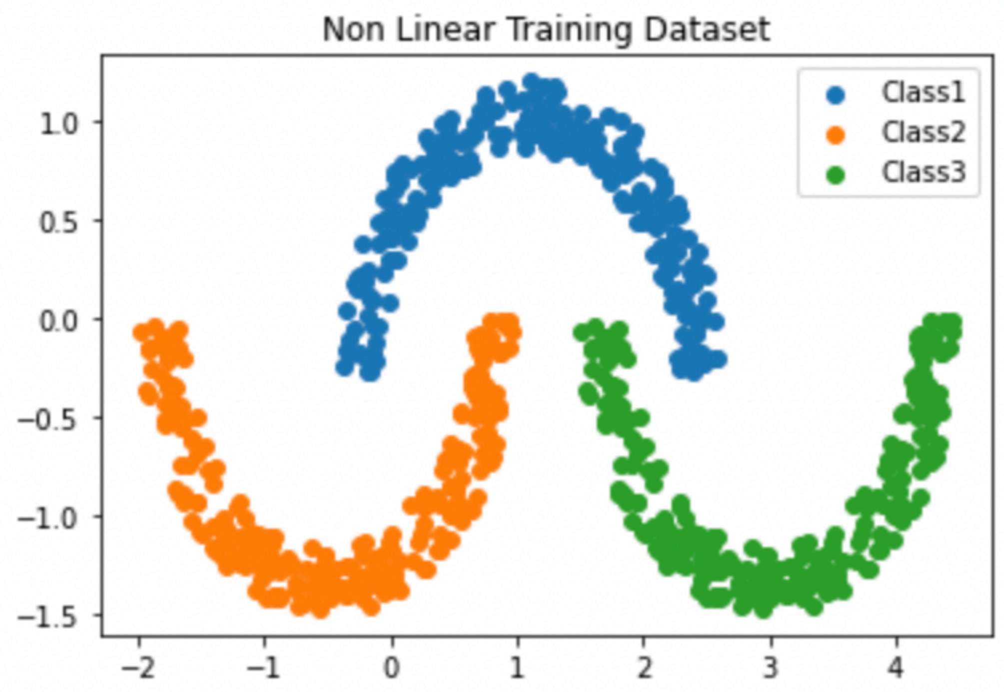


From the above ROC curve compared between two features STE and MEL, **MEL peafure performs better than STE feature**! for classification of audio into speech or non-speech.

BAYES CLASSIFIER (THREE CLASS CLASSIFICATION)

Given two datasets of three classes each. 50% training and 50% testing separated. One is linearly separable data and the other is non linearly separable data.





BAYES CLASSIFIER

The liklihood of the multimodal gaussian distribution is given below,

$$L(\mu, \sigma | x) = \frac{1}{2\pi\sqrt{|\sum|}} e^{-\frac{1}{2}((x-\mu)\sum^{-1}(x-\mu)^T}$$

Where μ is the mean of the samples and $\sum$ is the covariance of the training samples.

BAYES THEOREM & PREDICTION

$$p(c_1 | x_i) = \frac{p(x_i | c_1)p(c_1)}{p(x_i | c_1)p(c_1) + p(x_i | c_2)p(c_2) + p(x_i | c_3)p(c_3)}$$

$$p(c_2 | x_i) = \frac{p(x_i | c_2)p(c_2)}{p(x_i | c_1)p(c_1) + p(x_i | c_2)p(c_2) + p(x_i | c_3)p(c_3)}$$

$$p(c_3 | x_i) = \frac{p(x_i | c_3)p(c_3)}{p(x_i | c_1)p(c_1) + p(x_i | c_2)p(c_2) + p(x_i | c_3)p(c_3)}$$

$$prediction = argmax(p(c_1 | x_i), p(c_2 | x_i), p(c3 | x_i))$$

CLASSIFIER 1

Covariance for all classes is $I\sigma^2$. Using the average of the sample variances for all dimensions, for all classes, from the training data as σ^2 .

$$\sum_1 = \sum_2 = \sum_3 = \sum = I\sigma^2$$

CLASSIFIER 2

Full but equal covariance for all classes, $\sum$. Use the average of the sample covariance matrix from all classes in the train data as $\sum$.

$$\sum = \frac{\Sigma_1 + \Sigma_2 + \Sigma_3}{3}$$

CLASSIFIER 3

Diagonal covariance matrix, distinct for each class. Use variances from the sample covariance matrix for each class

$$\sum_1 = diagonal(\sigma_1^2)$$

$$\sum_2 = diagonal(\sigma_2^2)$$

$$\sum_3 = diagonal(\sigma_3^2)$$

CLASSIFIER 4

Full covariance matrix, distinct for each class. Use the sample covariance matrix for each class.

$$\sum_1, \sum_2, \sum_3$$

CONFUSION MATRIX

THE CONFUSION MATRIX

		ACTUAL CLASS		
Predicted Class		Class 1	Class 2	Class 3
	Class 1	C11	C12	C13
	Class 2	C21	C22	C23
	Class 3	C31	C32	C33

- **C11**: 98 test samples predicted as class 1 & actually belongs to class 1
- **C12**: 0 test samples predicted as class 1 & actually belongs to class 2
- **C13**: 2 test samples predicted as class 1 & actually belongs to class 3
- **C21**: 0 test samples predicted as class 2 & actually belongs to class 1
- **C22**: 100 test samples predicted as class 2 & actually belongs to class 2
- **C23**: 0 test samples predicted as class 2 & actually belongs to class 3
- **C31**: 0 test samples predicted as class 3 & actually belongs to class 1
- **C32**: 0 test samples predicted as class 3 & actually belongs to class 2
- **C33**: 100 test samples predicted as class 3 & actually belongs to class 3

ACCURACY

$$Accuracy = \frac{\text{Number of samples correctly classified (C11+C22+C33)}}{\text{Total number of samples used for testing}} * 100$$

MEAN PRECISION

"*Number of samples correctly classified as positive class, out of all the examples classified as positive class***"**

$$\text{Class 1 prescision} = \frac{\text{TP for Class 1}}{\text{TP for Class 1 + FP for class 1}}$$

$$\text{Class 2 prescision} = \frac{\text{TP for Class 2}}{\text{TP for Class 2 + FP for class 2}}$$

$$\text{Class 3 prescision} = \frac{\text{TP for Class 3}}{\text{TP for Class 3 + FP for class 3}}$$

$$\textit{Mean prescision} = \frac{1}{3}\sum_{i=1}^{3} Presision_i$$

RECALL

"*Number of samples correctly classified as positive class, out of all the examples belonging to positive class*"

$$\text{Class 1 Recall} = \frac{\text{TP for Class 1}}{\text{TP for Class 1 + FN for class 1}}$$

$$\text{Class 2 Recall} = \frac{\text{TP for Class 2}}{\text{TP for Class 2 + FP for class 2}}$$

$$\text{Class 3 Recall} = \frac{\text{TP for Class 3}}{\text{TP for Class 3 + FN for class 3}}$$

$$\textit{Mean recall} = \frac{1}{3}\sum_{i=1}^{3} Recall_i$$

F-MEASURE

"*Harmonic mean of precision and recall*"

$$\text{Class 1 F measure} = 2\frac{\text{Precision}_1 * \text{Recall}_1}{\text{Precision}_1 + \text{Recall}_1}$$

$$\text{Class 2 F measure} = 2\frac{\text{Precision}_2 * \text{Recall}_2}{\text{Precision}_2 + \text{Recall}_2}$$

$$\text{Class 3 F measure} = 2\frac{\text{Precision}_3 * \text{Recall}_3}{\text{Precision}_3 + \text{Recall}_3}$$

$$\textit{Mean Fmeasure} = \frac{1}{3}\sum_{i=1}^{3} Fmeasure_i$$

LINEARLY SEPARABLE (LS) DATA RESULTS

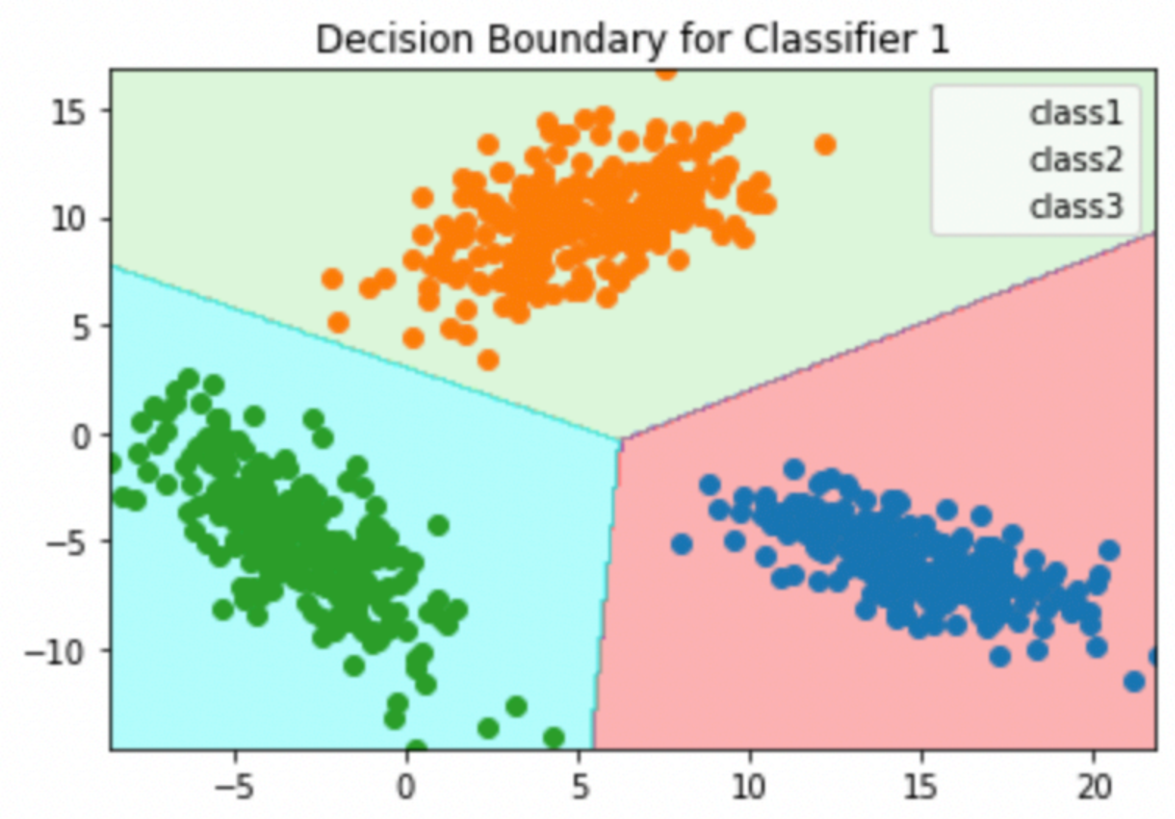
CLASSIFIER 1 RESULTS

CLASSIFIER 1 (LS)

ACTUAL CLASS				
Predicted Class		Class 1	Class 2	Class 3
	Class 1	250	0	0
	Class 2	0	250	0
	Class 3	0	0	250

SCORES CLASSIFIER 1 (LS)

	Classifier1
Accuracy	1.0
Precision	1.0
Recall	1.0
F Score	1.0



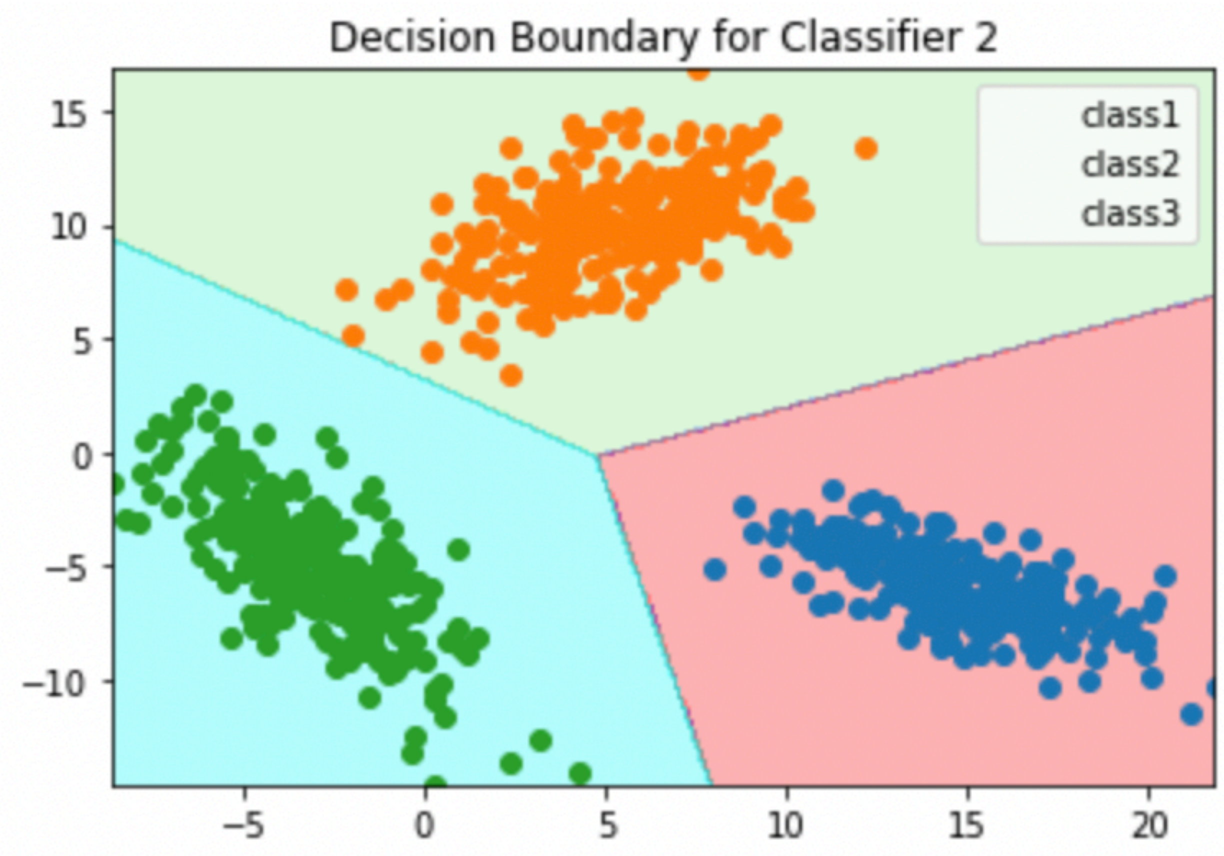
CLASSIFIER 2 RESULTS

CLASSIFIER 2 (LS)

		ACTUAL CLASS		
Predicted Class		Class 1	Class 2	Class 3
	Class 1	250	0	0
	Class 2	0	250	0
	Class 3	0	0	250

SCORES CLASSIFIER 2 (LS)

	Classifier1	Classifier2
Accuracy	1.0	1.0
Precision	1.0	1.0
Recall	1.0	1.0
F Score	1.0	1.0



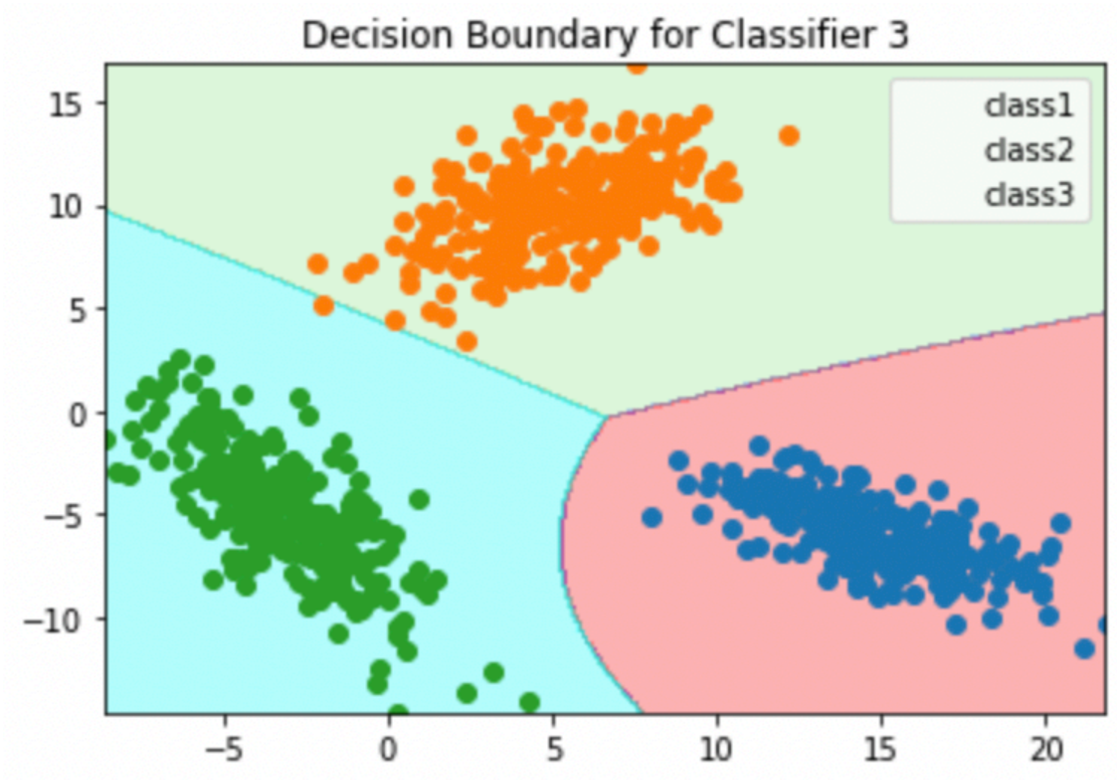
CLASSIFIER 3 RESULTS

CLASSIFIER 3 (LS)

	ACTUAL CLASS			
Predicted Class		Class 1	Class 2	Class 3
	Class 1	250	0	0
	Class 2	0	250	0
	Class 3	0	0	250

SCORES CLASSIFIER 3 (LS)

	Classifier1	Classifier2	Classifier3
Accuracy	1.0	1.0	1.0
Precision	1.0	1.0	1.0
Recall	1.0	1.0	1.0
F Score	1.0	1.0	1.0



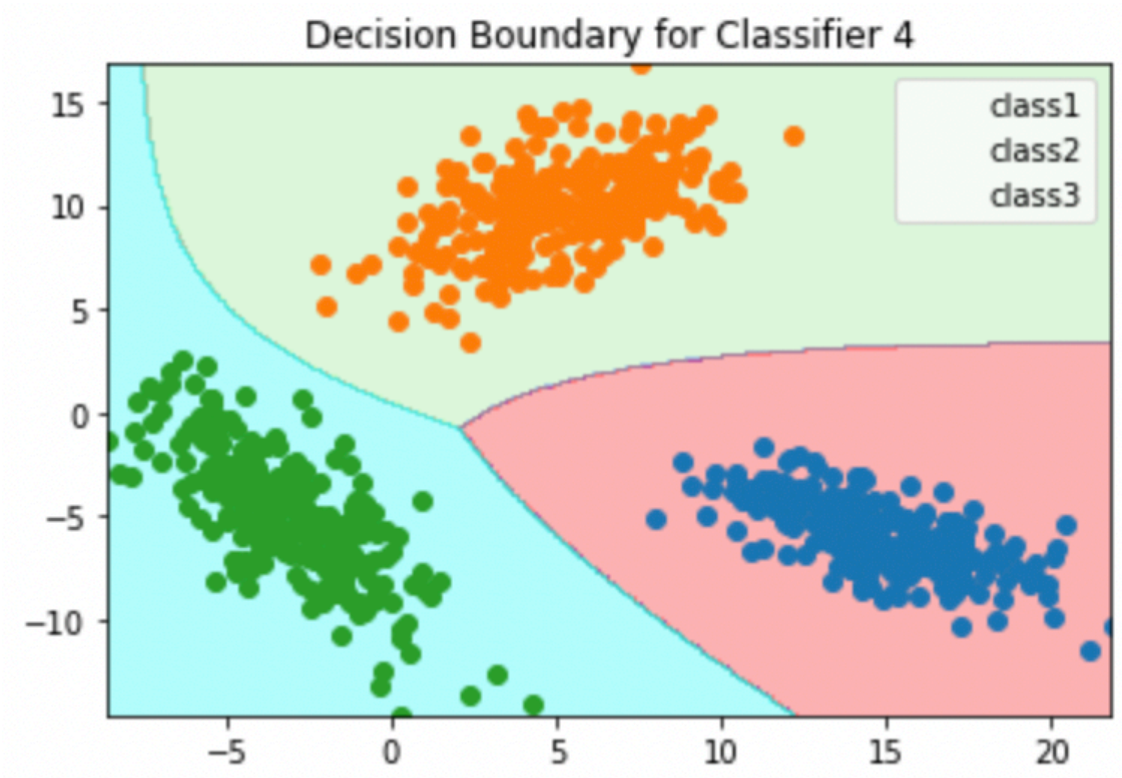
CLASSIFIER 4 RESULTS

CLASSIFIER 4 (LS)

	ACTUAL CLASS			
Predicted Class		Class 1	Class 2	Class 3
	Class 1	250	0	0
	Class 2	0	250	0
	Class 3	0	0	250

SCORES CLASSIFIER 4 (LS)

	Classifier1	Classifier2	Classifier3	Classifier4
Accuracy	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
Recall	1.0	1.0	1.0	1.0
F Score	1.0	1.0	1.0	1.0



CONCLUSION

SCORES CONCLUSION

	Classifier1	Classifier2	Classifier3	Classifier4
Accuracy	1.0	1.0	1.0	1.0
Precision	1.0	1.0	1.0	1.0
Recall	1.0	1.0	1.0	1.0
F Score	1.0	1.0	1.0	1.0

For this dataset all the classifier models performs well and gives same result. The differences we can note in the decision boundaries of each classifier.

As the covariance matrix becomes distinct and full, the boundary becomes much more accurate and exhibits non-linear characteristics.

NON LINEARLY SEPARABLE (NLS) DATA RESULTS

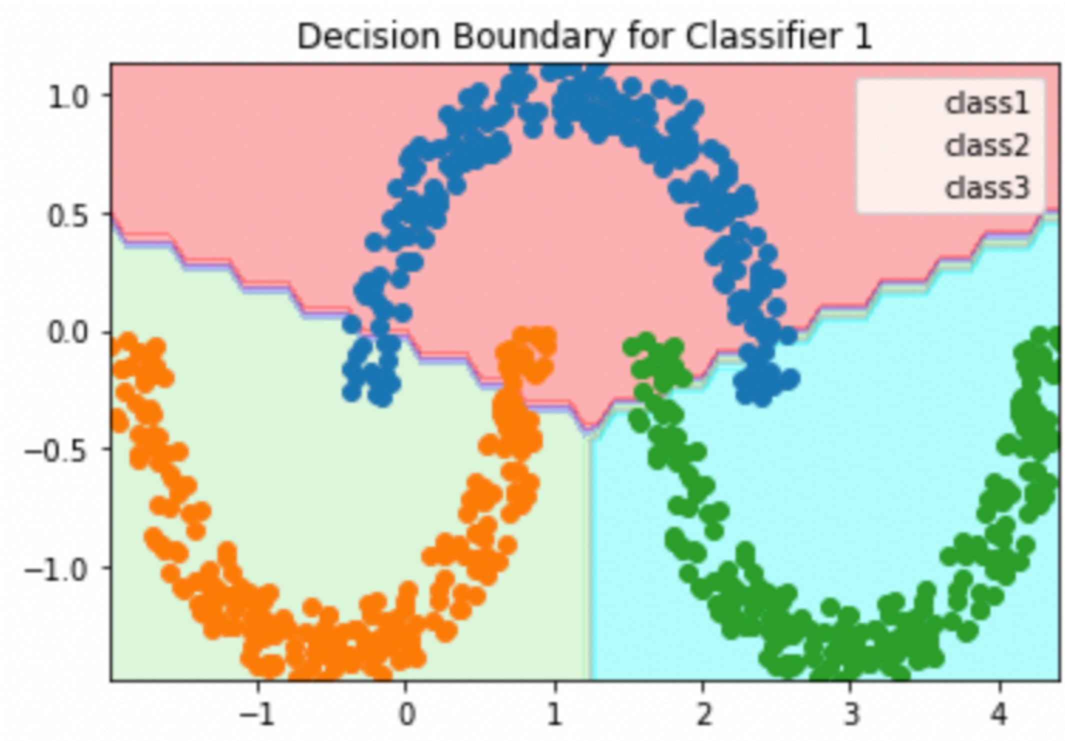
CLASSIFIER 1 RESULTS

CLASSIFIER 1 (NLS)

	ACTUAL CLASS			
		Class 1	Class 2	Class 3
Predicted Class	Class 1	222	16	22
	Class 2	0	234	0
	Class 3	0	0	228

SCORES CLASSIFIER 1 (NLS)

	Classifier1
Accuracy	0.912
Precision	0.951
Recall	0.949
F Score	0.947



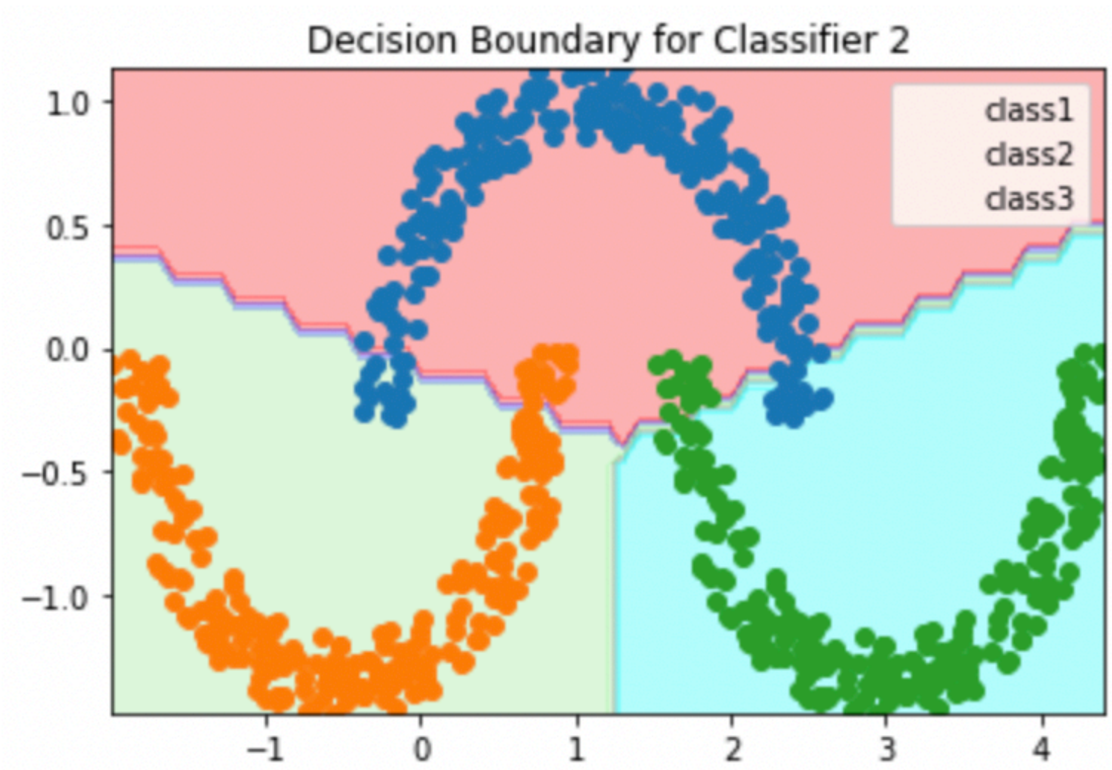
CLASSIFIER 2 RESULTS

CLASSIFIER 2 (NLS)

	ACTUAL CLASS			
Predicted Class		Class 1	Class 2	Class 3
	Class 1	222	16	22
	Class 2	0	234	0
	Class 3	0	0	228

SCORES CLASSIFIER 2 (NLS)

	Classifier1	Classifier2
Accuracy	0.912	0.912
Precision	0.951	0.951
Recall	0.949	0.949
F Score	0.947	0.947



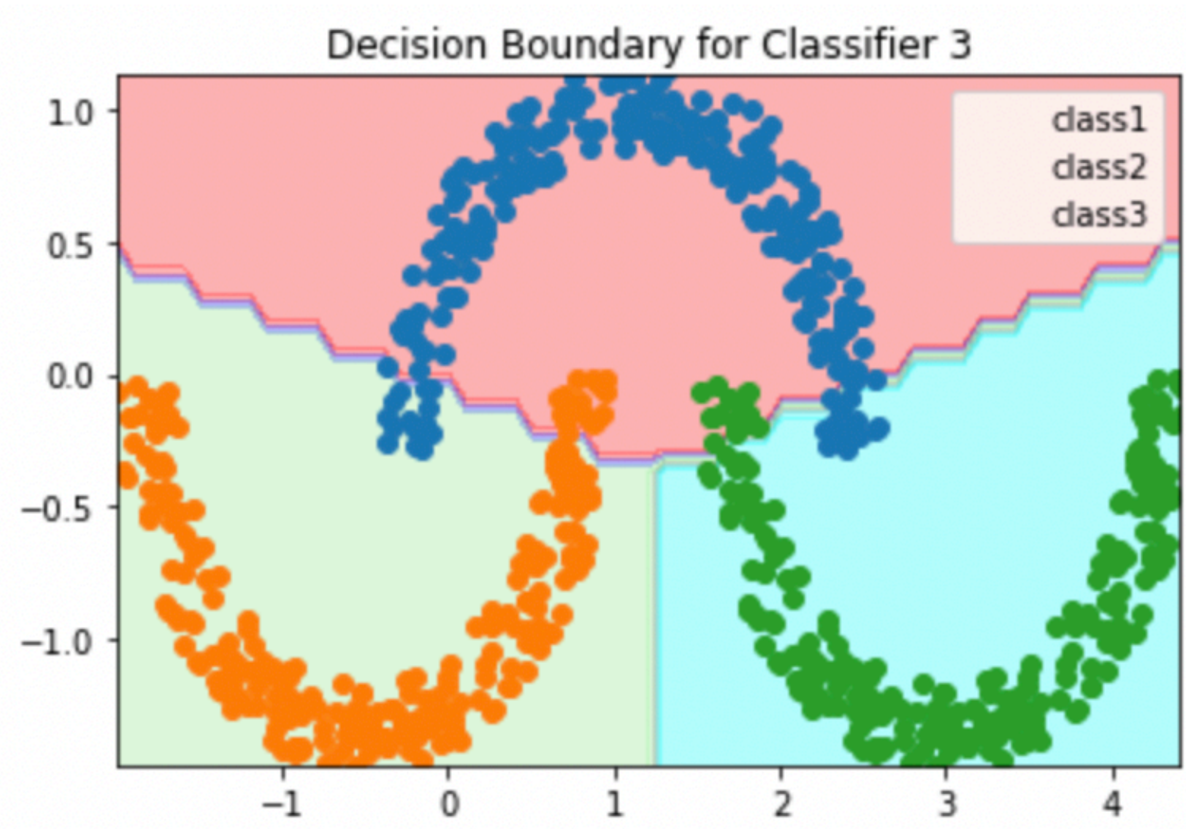
CLASSIFIER 3 RESULTS

CLASSIFIER 3 (NLS)

ACTUAL CLASS				
		Class 1	Class 2	Class 3
Predicted Class	Class 1	220	12	21
	Class 2	0	238	0
	Class 3	0	0	229

SCORES CLASSIFIER 3 (NLS)

	Classifier1	Classifier2	Classifier3
Accuracy	0.912	0.912	0.916
Precision	0.951	0.951	0.956
Recall	0.949	0.949	0.956
F Score	0.947	0.947	0.953



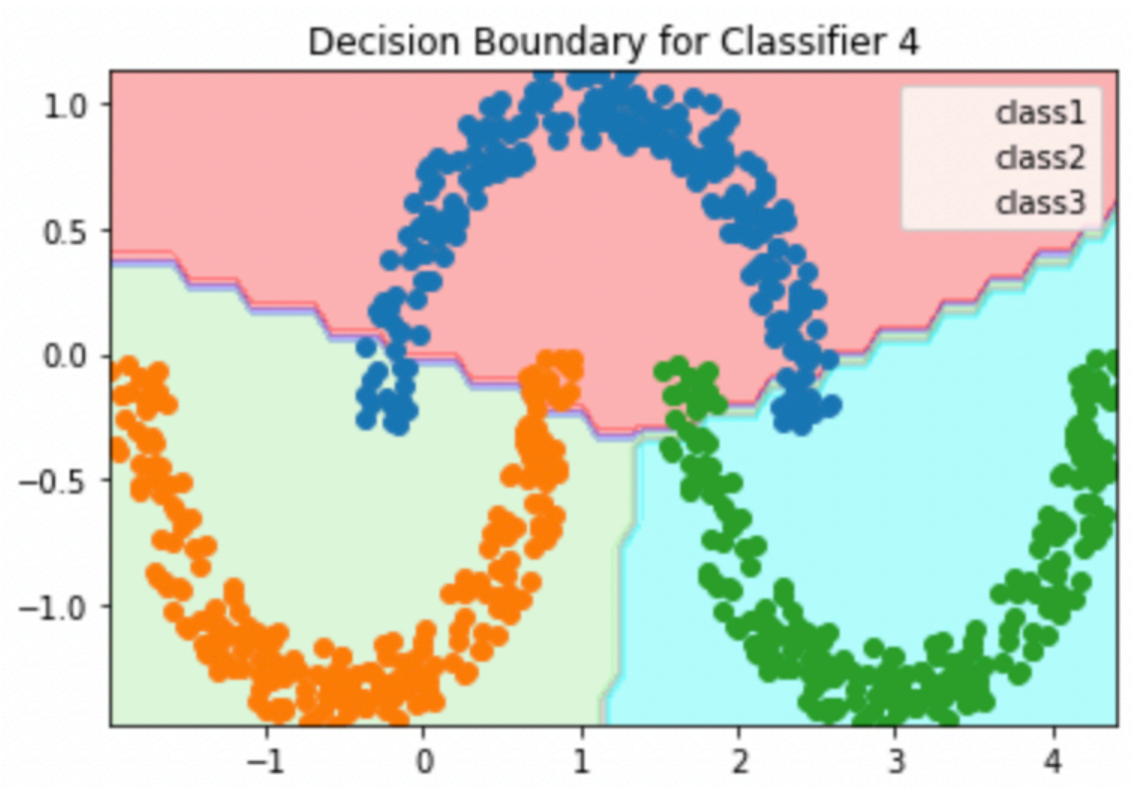
CLASSIFIER 4 RESULTS

CLASSIFIER 4 (NLS)

	ACTUAL CLASS			
		Class 1	Class 2	Class 3
Predicted Class	Class 1	221	9	22
	Class 2	0	241	0
	Class 3	0	0	228

SCORES CLASSIFIER 4 (NLS)

	Classifier1	Classifier2	Classifier3	Classifier4
Accuracy	0.912	0.912	0.916	0.92
Precision	0.951	0.951	0.956	0.958
Recall	0.949	0.949	0.956	0.958
F Score	0.947	0.947	0.953	0.956



CONCLUSION

SCORES CONCLUSION (NLS)

	Classifier1	Classifier2	Classifier3	Classifier4
Accuracy	0.912	0.912	0.916	0.92
Precision	0.951	0.951	0.956	0.958
Recall	0.949	0.949	0.956	0.958
F Score	0.947	0.947	0.953	0.956

For this non linear dataset all the classifier models does not performs very well, but gives some reasonable result. The differences we can note in the decision boundaries of each classifier also in the scores of the classifiers.

As the covariance matrix becomes distinct and full, the boundary becomes much more accurate and the accuracy, precision, recall, and the Fscore also imporves.

DETAILS & LINKS OF CODE

Name	**Rajesh R**
Roll No	**S21005**
Mail	**s21005@students.iitmandi.ac.in**
GitHub	**https://github.com/its-rajesh/Pattern-Recognition**
Colab Question1	**https://colab.research.google.com/drive/1-sa7xiG_oQq43cVfhDanlptJ2-GVjzvc?usp=sharing**
Colab Question2	**https://colab.research.google.com/drive/1bv4QfTNr_DmIs3bLR5XMHg7Sa3iRi-KM?usp=sharing**