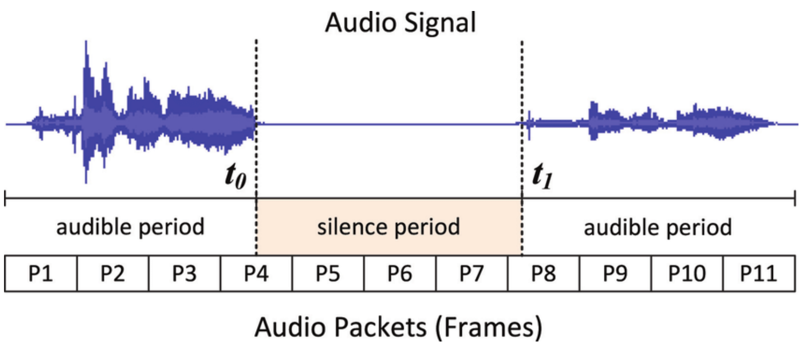
PATTERN RECOGNITION ASSIGNMENT-1

BY-VIKAS SINGH(T21017)

10/10/2022

1 PROBLEM-1:

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



MAXIMUM LIKELIHOOD FOR NORMAL DISTRIBUTION

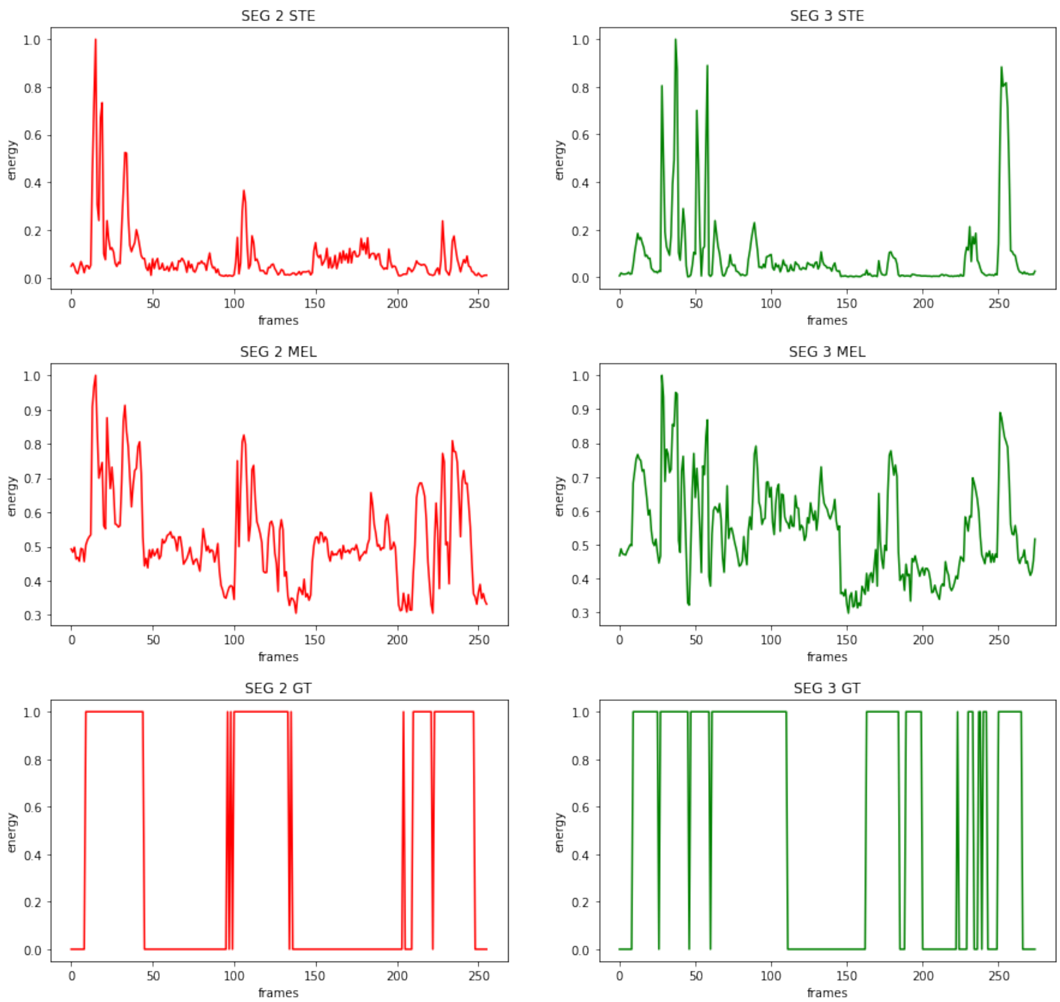
Gaussian probability distribution is perhaps the most used distribution in all of science.
A simple unimodal Gaussian to estimate the distribution of the features. The task is to find the mean and variance for estimation. The liklihood function is given by,

$P(x) = (1/\sqrt{2\pi}\sigma) * \exp(-1/2((x-\mu)/\sigma)^2)$

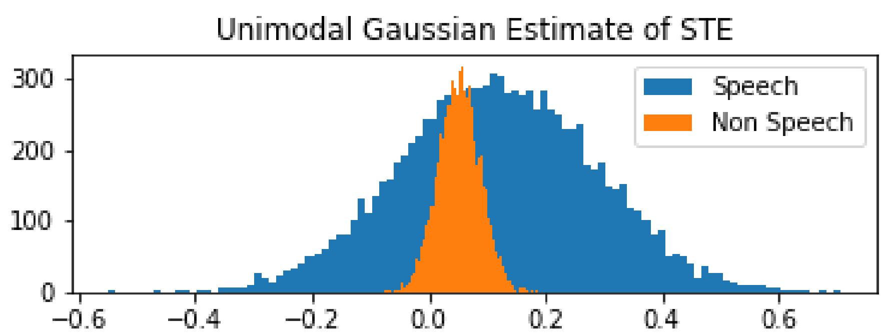
Mean -
$\mu = (x_1 + x_2 + ... + x_n)/n$
Var-
$\sigma^2 = ((x_1 - \mu)^2 + (x_2 - \mu)^2 + ...)/n$

After estimating the mean and variance of the speech and non-speech dataset, we can fix the distribution for further calculations for classifications.

STE,MEL,GT FEATURES FOR TRAINING AND TEST DATASET



STE feature



Classification

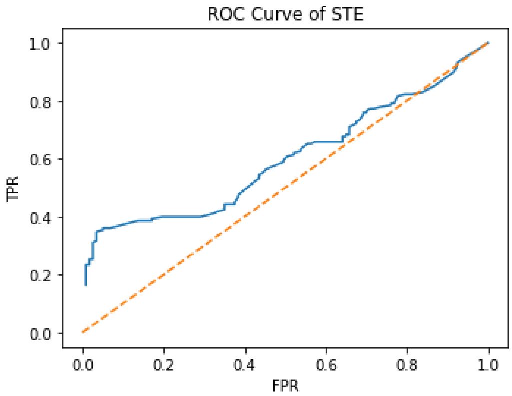
Here the classification of speech or non-speech is based on the posteriori which is given by-

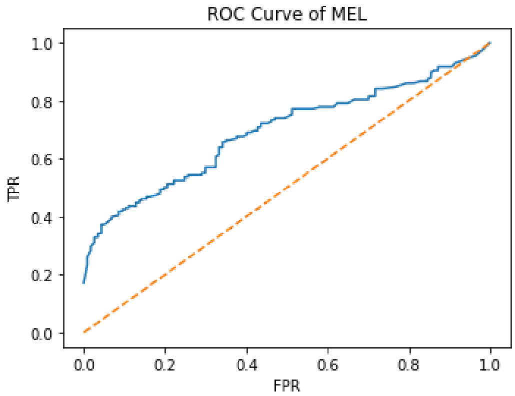
$$p(s/x_i) = (p(x_i/s) * p(s))/p(x_i)$$

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

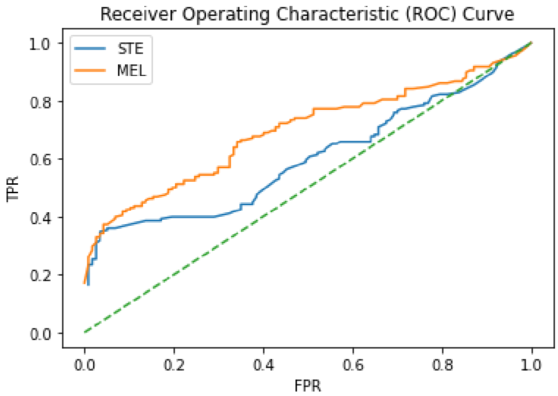
ROC

A ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR). The true positive rate is the proportion of observations that were correctly predicted to be positive out of all positive observations (TP/(TP + FN)). Similarly, the false positive rate is the proportion of observations that are incorrectly predicted to be positive out of all negative observations (FP/(TN + FP))
A discrete classifier that returns only the predicted class gives a single point on the ROC space. But for probabilistic classifiers, which give a probability or score that reflects the degree to which an instance belongs to one class rather than another, we can create a curve by varying the threshold for the score. Note that many discrete classifiers can be converted to a scoring classifier by looking inside their instance statistics





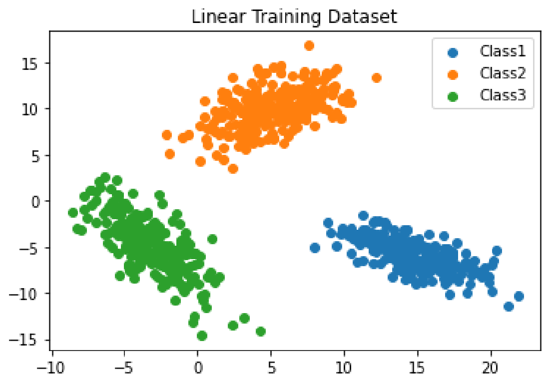
Comparision of ROCs

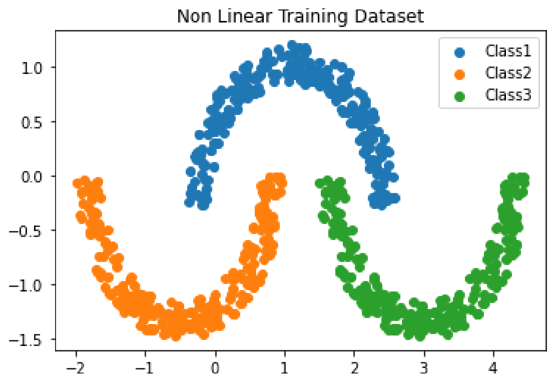


Because MEL has more area hence MEL is better than STE for classification of audio into speech or non-speech.

2 PROBLEM 2

input data





BAYES CLASSIFIER

The liklihood of the multimodal gaussian distribution is given below,

$p(x) = 1/(2\pi\sqrt{|cov|}) * exp(-0.5 * (x - \mu) * cov^{-1} * (x - \mu).T)$

Where μ is the mean of the samples and cov is the covariance of the training samples.

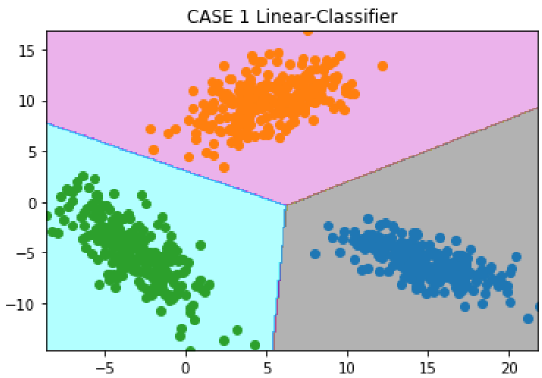
Bayes theorem and prediction

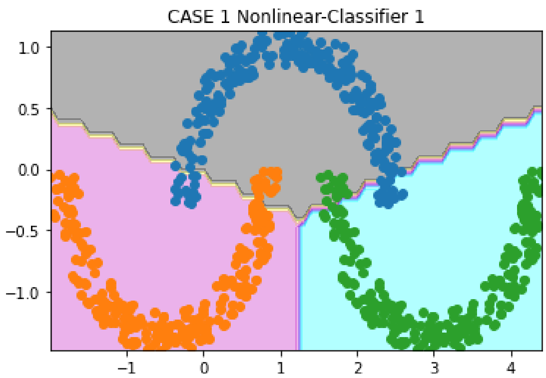
$p(c_1/x_i) = (p(x_i)/p(c_1))p(c_1)/(p(x_i/(c_1))p(c_1)+p(x_i/(c_2))p(c_2)+p(x_i/c_3))p(c_3)$
$p(c_2/x_i) = (p(x_i)/p(c_2))p(c_2)/(p(x_i/(c_1))p(c_1)+p(x_i/(c_2))p(c_2)+p(x_i/c_3))p(c_3)$
$p(c_3/x_i) = (p(x_i)/p(c_3))p(c_3)/(p(x_i/(c_1))p(c_1)+p(x_i/(c_2))p(c_2)+p(x_i/c_3))p(c_3)$

Prediction=argmax((c_1/x_i),(c_2/x_i),(c_3/x_i))

Classifier 1

Covariance for all classes is . Using the average of the sample variances for all dimensions, for all classes, from the training data as .
cov=cov1=cov2=cov3=$\sigma^2 I$

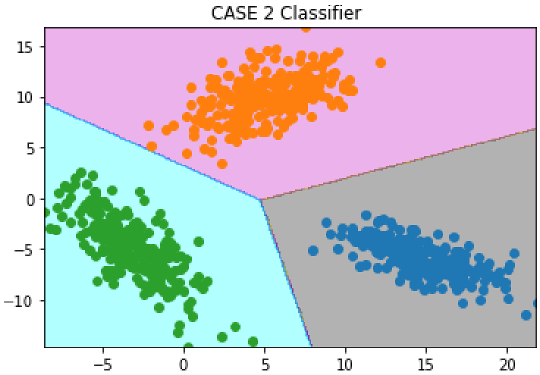


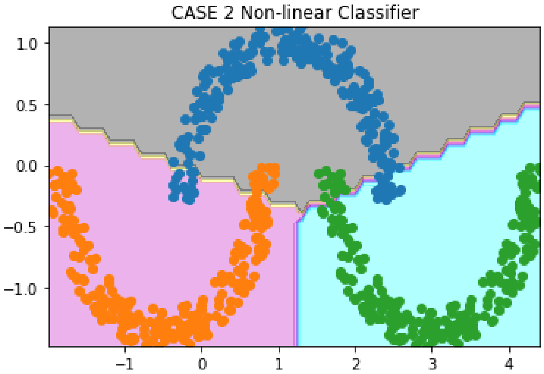


Classifier 2

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

cov=(cov1+cov2+cov3)/3

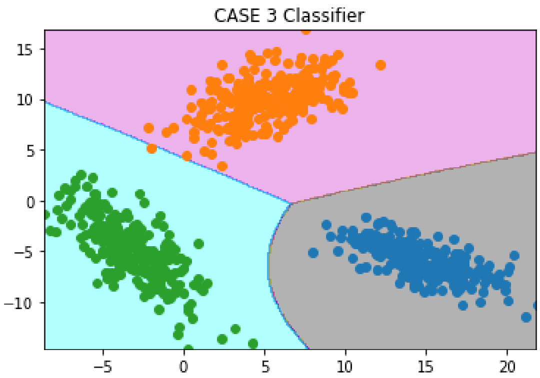


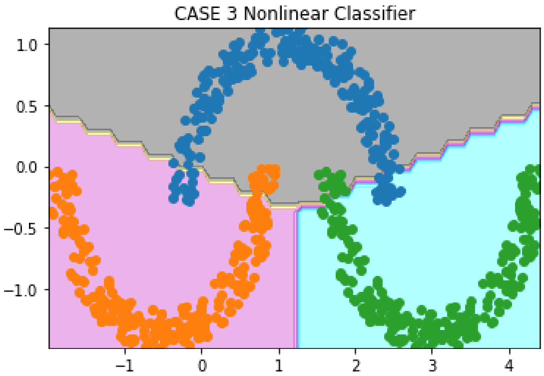


Classifier 3

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

cov1=digonal(σ_1^2)
cov2=digonal(σ_2^2)
cov3=digonal(σ_3^2)

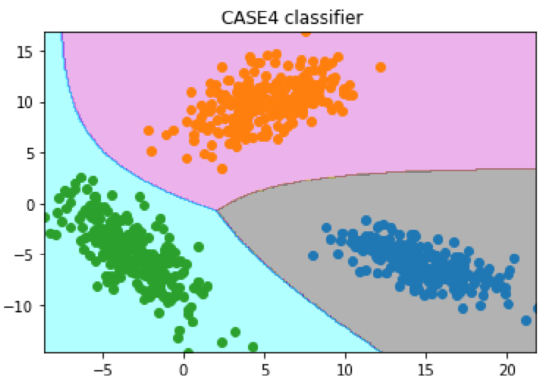


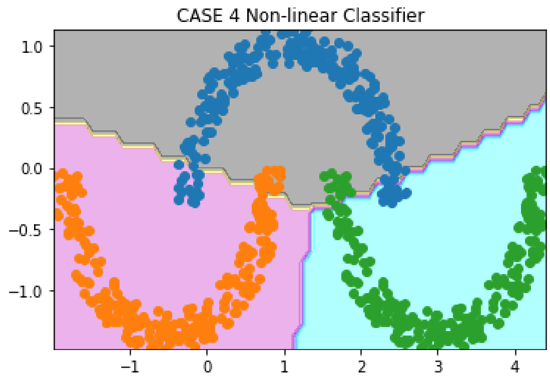


Classifier 4

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

cov1,cov2,cov3





2.1 Conclusion

Table

Index	Accuracy	Precision	Recall	F1-score
linear CASE1	1	1	1	1
nonlinear CASE1	0.912	0.9512	0.9493	0.9473
linear CASE2	1	1	1	1
nonlinear CASE2	0.912	0.9512	0.9493	0.9473
linear CASE3	1	1	1	1
nonlinear CASE3	0.916	0.9565	0.9565	0.9539
linear CASE4	1	1	1	1
nonlinear CASE4	0.912	0.9590	0.9587	0.9567

For linear 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. For 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 improves