Department of Computer Engineering

University of Peradeniya

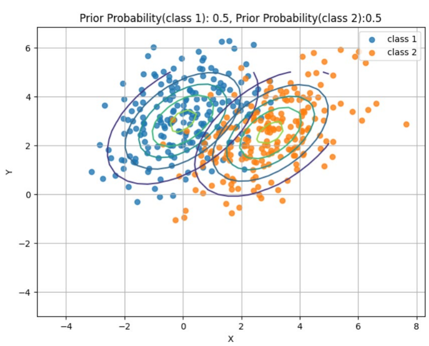
Machine Learning Lab Three

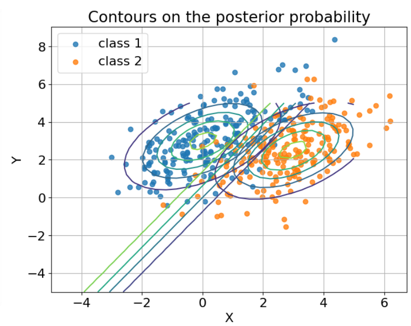
E/19/306

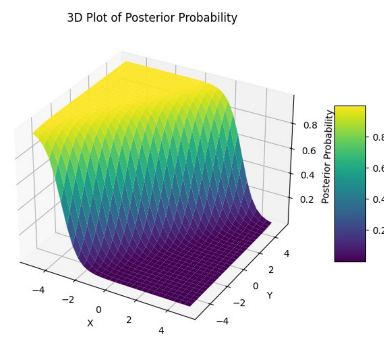
Rajakaruna M.M.P.N.

1 Class Boundaries and Posterior Probabilities

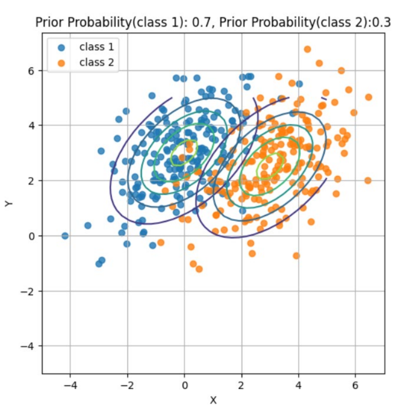
$\boldsymbol{m}_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$ and $\boldsymbol{m}_1 = \begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$, $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $P_1 = P2 = 0.5$;

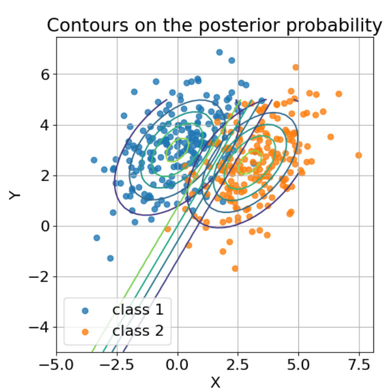


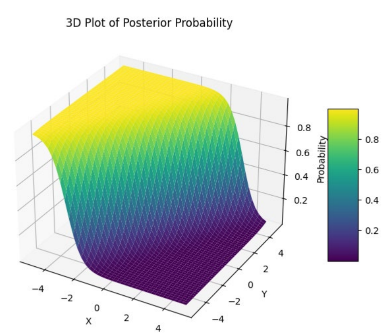




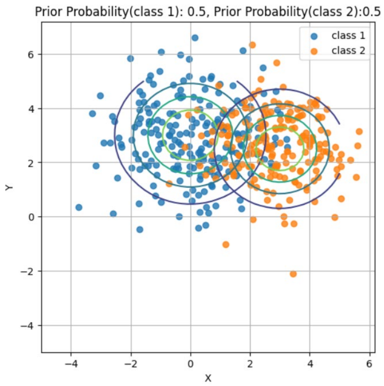
$\boldsymbol{m}_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$ and $\boldsymbol{m}_2 = \begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$, $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $P_1 = 0.7,\ P_2 = 0.3$;

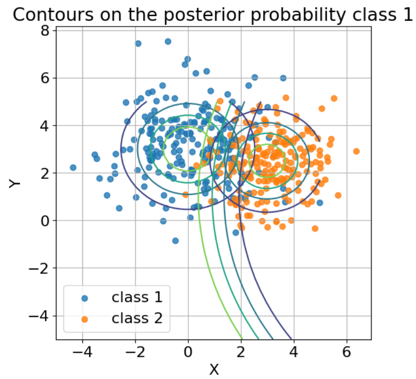


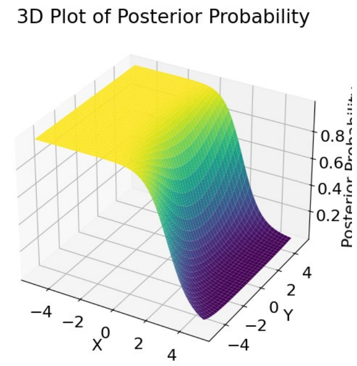




$\boldsymbol{m}_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$ and $\boldsymbol{m}_2 = \begin{pmatrix} 3 \\ 2.5 \end{pmatrix}$, $\boldsymbol{C}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\boldsymbol{C}_2 = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$, $P_1 = P2 = 0.5$.







From Bayes theorem the posterior probability for class C1 can be written as,

$p(\mathcal{C}_1|\mathbf{x}) \quad = \quad \frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)}$ If these are normally distributed, and we divide by the numerator we find that the posterior probability surface is actually of sigmoidal shape.

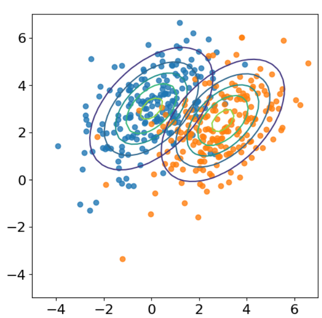
$\sigma(a) = \frac{1}{1+\exp(-a)} \qquad a = \ln\frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)}$, Let us assume that the class-conditional densities are Gaussian and then explore the resulting form for the posterior probabilities.

$p(\mathbf{x}|\mathcal{C}_k) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\mathbf{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right\}$. Therefore ‘a’ can be written as $a = -\frac{1}{2}\,(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) + \frac{1}{2}\,(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + ln\frac{p(C_2)}{p(C_1)}$

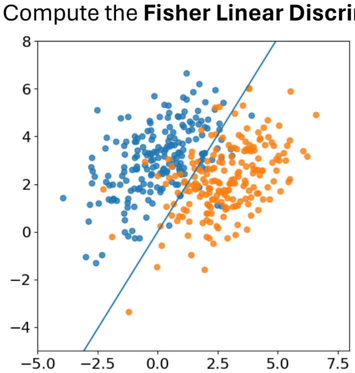
2 Fisher LDA and ROC Curve

Plot contours on the two densities.

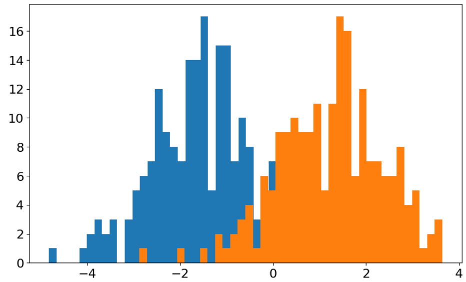
Draw 200 samples from each of the two distributions and plot them on top of the contours.



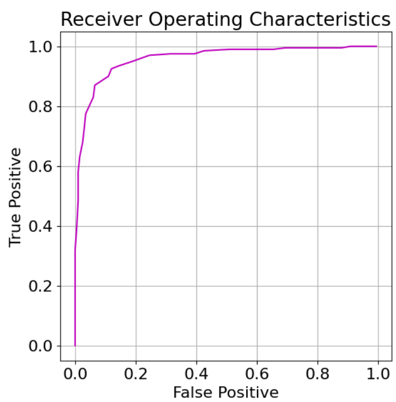
Compute the **Fisher Linear Discriminant** direction and plot. wF = (C1 + C2)−1 (m1 − m2)



Project the data onto the Fisher discriminant directions and plot histograms of the distribution of projections.



Compute and plot the Receiver Operating Characteristic (ROC) curve.

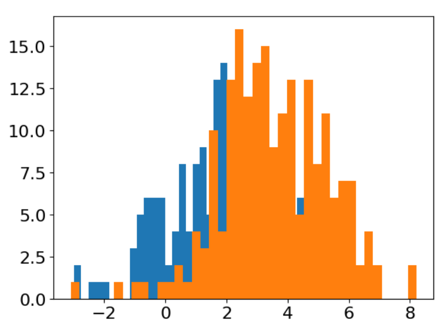


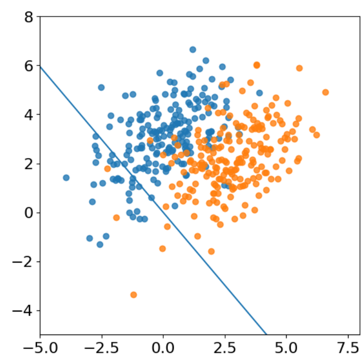
Area under curve (AUC): 0.9541875

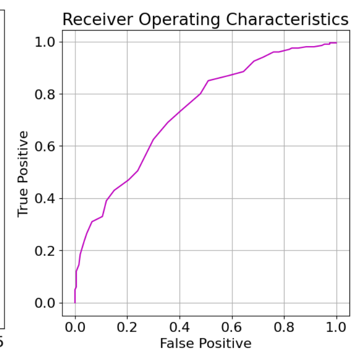
For a suitable choice of decision threshold, compute the classification accuracy

Best Threshold: 1.0187773004750094   Max Accuracy: 0.8825000000000001

Projecting Classes onto a Random Direction

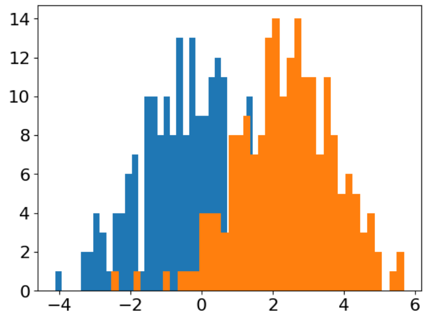


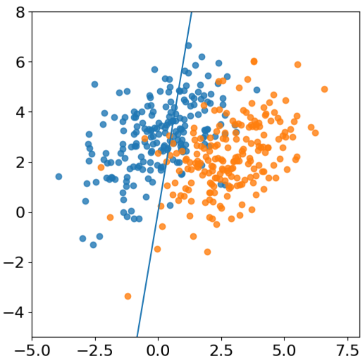


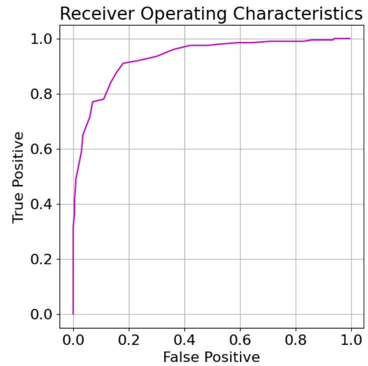


Area under curve (AUC): 0.7326874999999999

For direction connecting the means

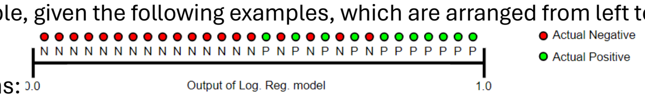






Area under curve (AUC): 0.9264125000000001

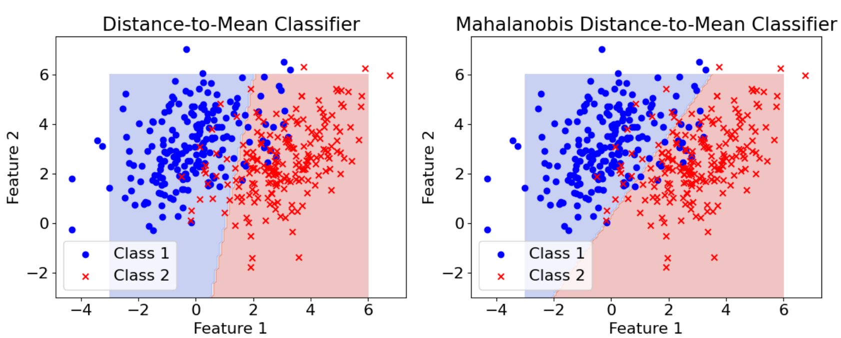
What is Area Under Curve (AUC)?

AUC stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve. AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. For example, given the following examples, which are arranged from left to right in ascending order of logistic regression predictions: 

AUC represents the probability that a random positive (green) example is positioned to the right of a random negative (red) example. AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0

By that we can say logistic regression that used the Fisher discriminant is a good way (AOC=0.9541875)

3 Mahalanobis Distance



In the distance-to-mean classifier, the decision boundary is simply a straight line perpendicular to the line connecting the two class means.

On the other hand, the Mahalanobis distance-to-mean classifier considers the covariance matrices of the classes. The decision boundary is no longer linear but follows the shape of the Mahalanobis distance contours. This classifier takes into account the correlation between features and the scaling of each feature, leading to more flexible and accurate classification, especially when classes have different covariances or non-spherical shapes.