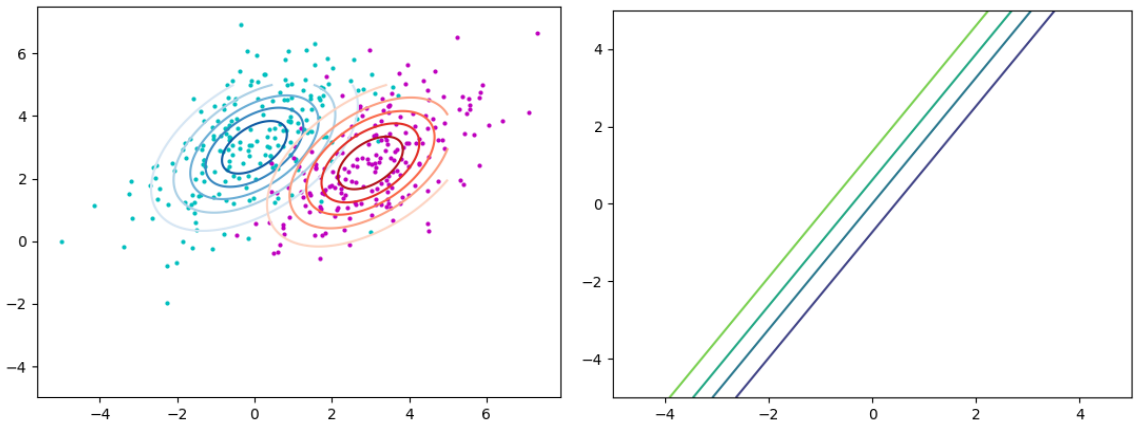
CO544: Machine Learning and Data Mining
Machine Learning Lab Three

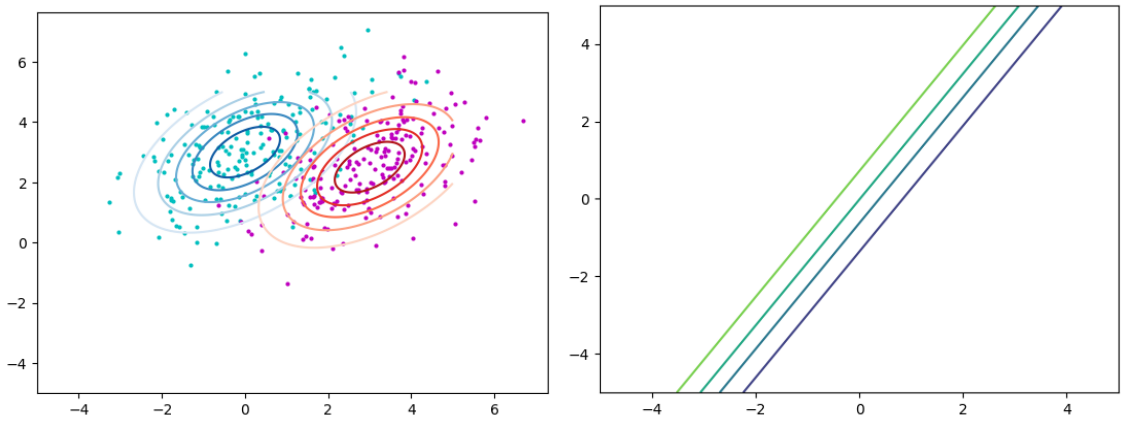
Ranage R.D.P.R. - E/19/310

1. Class Boundaries and Posterior Probabilities

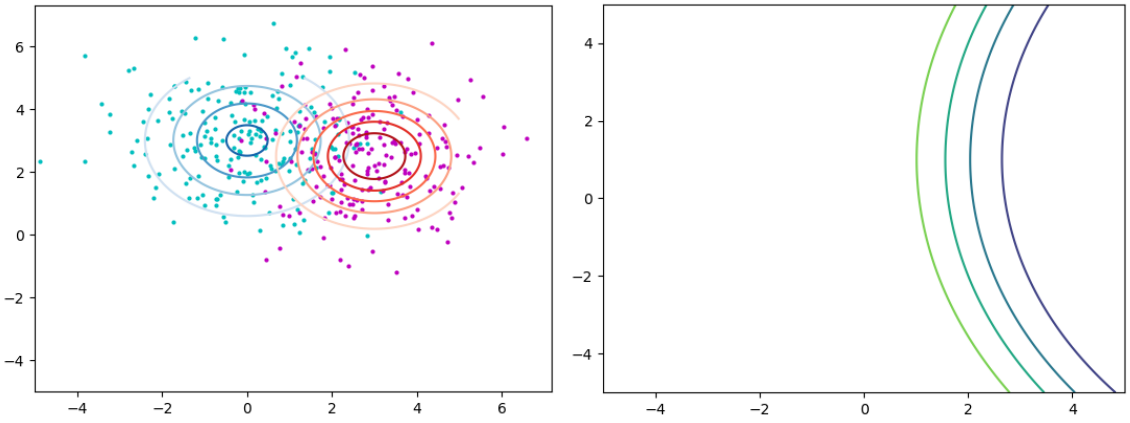
a) m1 = [[0, 3]],m2 = [[3, 2.5]] ,C1,C2 = [[2, 1], [1, 2]], P1,P2 = 0.5



b) m1 = [[0, 3]],m2 = [[3, 2.5]] ,C1,C2 = [[2, 1], [1, 2]], P1 = 0.7 , P2 = 0.3



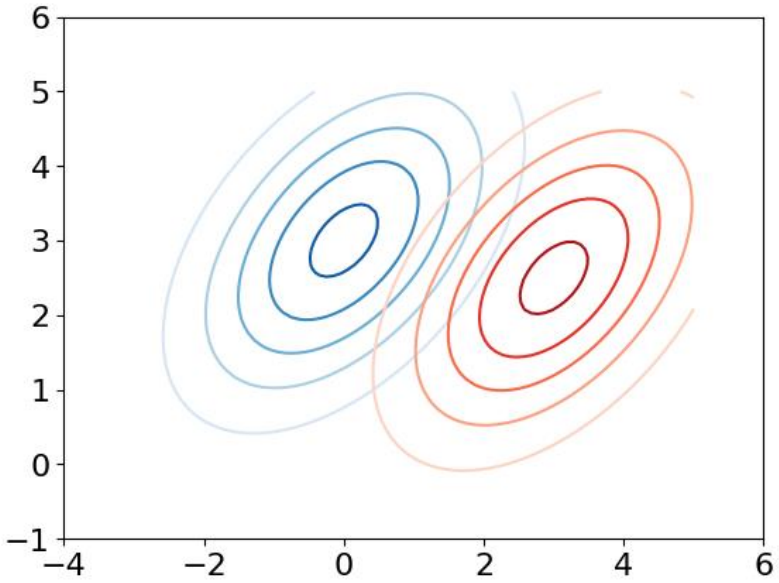
c) m1 = [[0, 3]],m2 = [[3, 2.5]] ,C1= [[2, 0], [0, 2]], C2 = [[1.5, 0], [0, 1.5]],, P1,P2 = 0.5



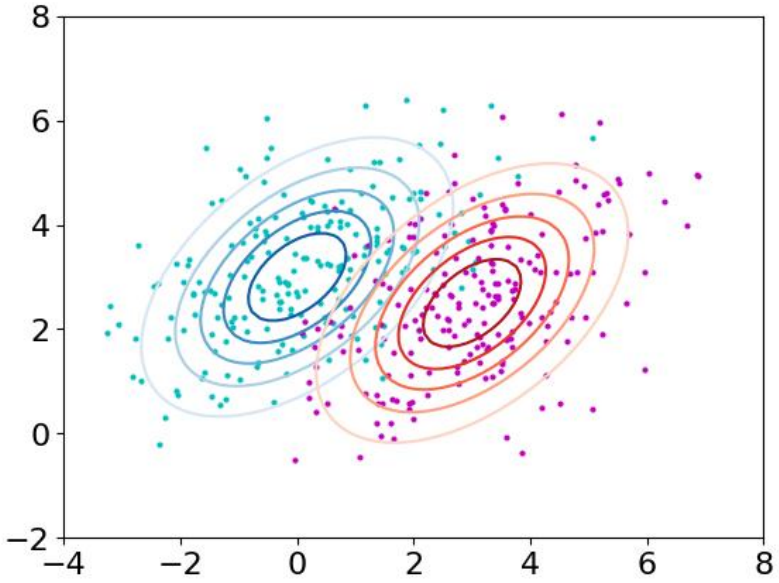
The likelihood contours in (a) are linear since both covariance matrices are the same. The class border is more likely to fall between the average of the two classes in this scenario since the prior probabilities are equal.
The likelihood contours in (b) are linear as both covariance matrices are the same. However, the prior probabilities—0.7 and 0.3—do not equal one another. The result is a movement in the class border in favor of the class with the greater prior probability (0.7).
In (c), the contour lines are nonlinear because the covariance matrices differ from one another. Given that the prior probabilities are identical, the class border is also more likely to fall between the means of the two classes.

2. Fisher LDA and ROC Curve

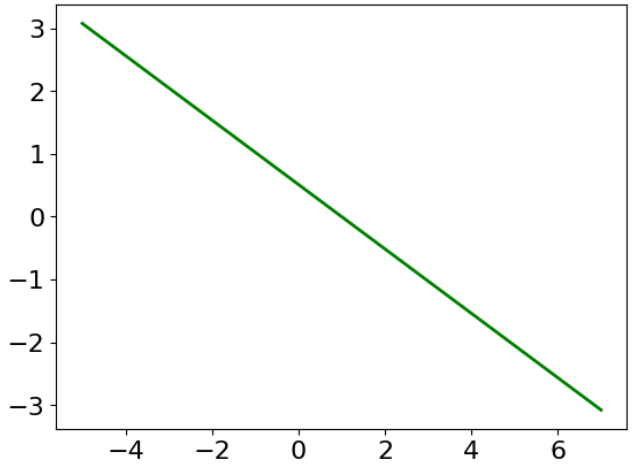
1. Plot contours on the two densities.



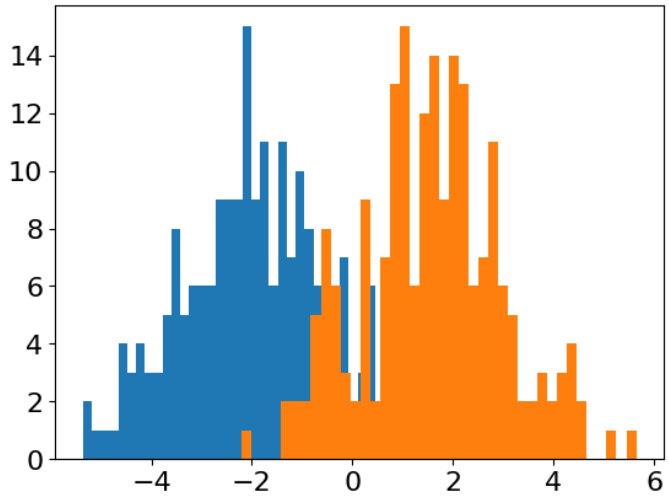
2. Draw 200 samples from each of the two distributions and plot them on top of the contours.



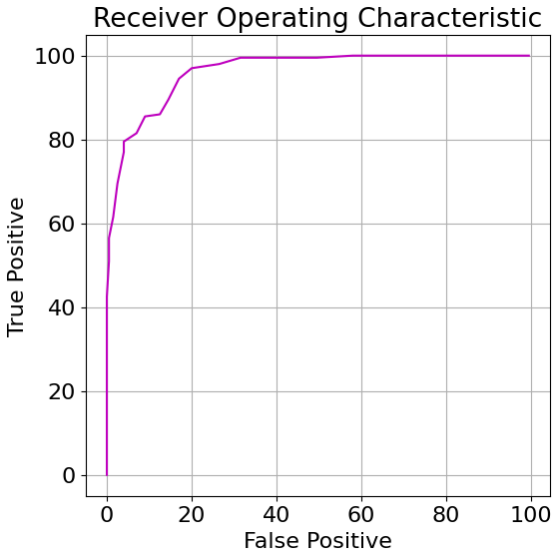
3. Compute the Fisher Linear Discriminant direction using the means and covariance matrices of the problem, and plot the discriminant direction: $wF = (C1 + C2)^{-1} (m1 - m2)$



4. Project the data onto the Fisher discriminant directions and plot histograms of the distribution of projections (an example of this is in Fig. 2(a);



5. Compute and plot the Receiver Operating Characteristic (ROC) curve, by sliding a decision threshold, and computing the True Positive and False Positive rates (see code snippet in Appendix and example of an ROC curve in Fig. 2(b).



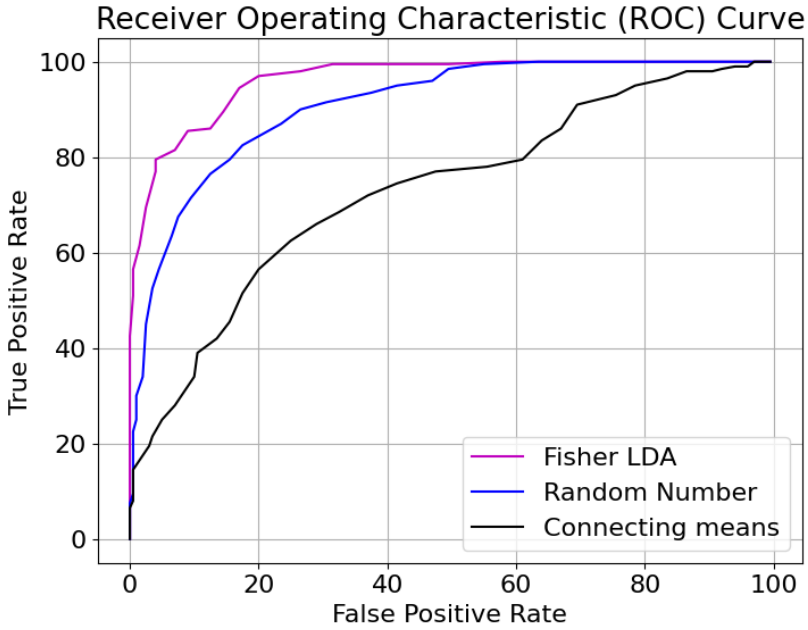
6. Compute the area under the ROC curve (Hint: try numpy.trapz)
Area under the curve : 9575.5

7. For a suitable choice of decision threshold, compute the classification accuracy.
Accuracy : 88.75
Max threshold: -0.64732441788028

8. Plot the ROC curve (on the same scale) for

- A random direction (instead of the Fisher discriminant direction).
- Projections onto the direction connecting the means of the two classes.

Compute the area under the ROC curve (AUC) for these two cases.



3. Mahalanobis Distance

Distance-to-Mean Classifier: In this classifier, we compute the Euclidean distance between the data point and the mean of each class. The class with the nearest mean is assigned to the data point. This classifier assumes that the classes have spherical Gaussian distributions with equal variances in all dimensions.

Mahalanobis Distance-to-Mean Classifier: This classifier accounts for different variances and covariances in each dimension by using the Mahalanobis distance. The Mahalanobis distance measures the distance between a point and a distribution, taking into account the covariance structure of the data. It is defined as the Euclidean distance after applying a linear transformation to the data to make the covariance matrix diagonal and equal to the identity matrix.

Overall, Mahalanobis distance to mean classifier performs somewhat better than the Distance to Mean classifier despite being more straightforward and computationally efficient due to its consideration of covariance structure and distribution means.Thus, for complex distributions, it is superior.