classifier affect the resulting decision boundary?

(10) 6. Describe the factors you would consider when deciding how to model the data (what assumptions to make regarding the covariance structure) for a Bayes classifier, and how you expect your classifier design choices (modeling assumptions) to affect the resulting classifier.

(5) 7. Submit a PDF print-out of your code for this section (Exploring Bayes Classifiers). (Submitting a URL for a cloud-based repository is insufficient.)

Comparing Linear Discriminant and Logistic Discriminant (and Bayes)

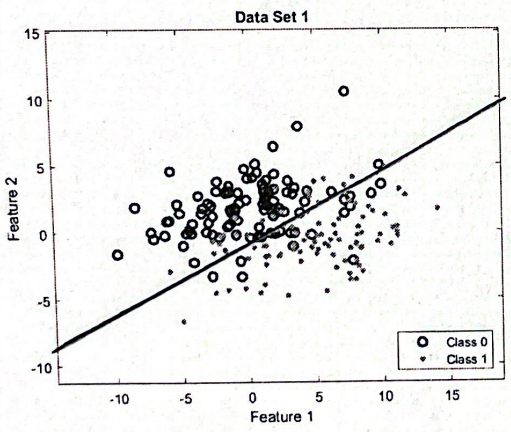
Linear Discriminant and Logistic Discriminant both assume a linear boundary separates the two classes; they differ in the assumptions they make to arrive at the resulting linear boundary and in how the coefficients for that boundary are computed. Here, you are going to explore how linear discriminant and logistic discriminant behave when operating on data that meet their underlying assumptions to different degrees. We will compare these linear classifiers to the Bayes classifier, to evaluate the implications of the linear boundary assumption.

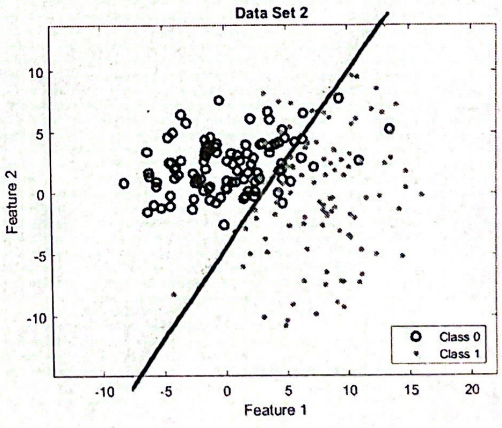
Make sure you are able to apply both a linear discriminant classifier and a logistic discriminant classifier.

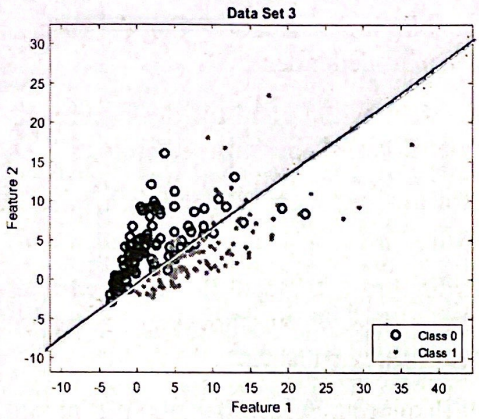
Regardless of whether you choose to write your own functions or leverage functions that may be available through Matlab or Python packages or libraries, you are responsible for understanding how the function(s) you are using work so you can effectively apply them to suit your needs and correctly interpret the results they provide.

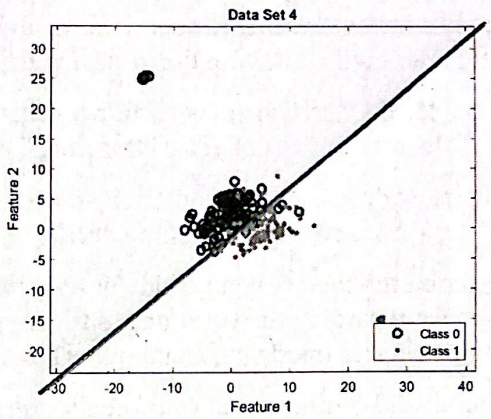
The following questions concern four data sets provided as a `csv` files:
`dataSet1.csv`, `dataSet2.csv`, `dataSet3.csv`, and `dataSet4.csv`.

Each csv file is organized such that each row contains the true class (either 0 or 1), followed by the associated (2-dimensional) feature vector. When you visualize the data sets, you should see this:









(5) 8. From visual inspection of these datasets (figures above), qualitatively sketch on the provided figures what you would consider to be a "good" linear decision boundary for each dataset (assuming the goal is $\max P_{cd}$ (or $\min P_e$) – the linear boundary you would draw if someone asked you to define the boundary.

(10) 9. Data set 1 is consistent with the assumptions underlying LDA – the data is Gaussian with means for the two classes that are distinct, and identical covariances.

(a) Apply the linear discriminant to dataset 1, and plot the decision statistic surface with both the training data and the decision boundary corresponding to $\lambda(x) = 0$ superimposed on top.

(b) Apply the logistic discriminant to dataset 1, and plot the decision statistic surface with both the training data and the decision boundary corresponding to $\lambda(x) = 0.5$ superimposed on top.

(c) Apply a Bayes Classifier to dataset 1, assuming the features may be dependent and the covariance matrices for the two classes are distinct (*i.e.*, estimate full covariance matrices for both class 0 and class 1), and plot the decision statistic surface for the ln-likelihood ratio with both the training data and the decision boundary under the assumptions of equal class priors and symmetric costs ($\ln \lambda(x) = 0$) superimposed on top.

(d) Compare the three decision boundaries (linear discriminant, logistic discriminant, and Bayes) by visualizing the data (replicating the figure provided for dataset 1 at the beginning of this section), and superimposing all three decision boundaries on top of the data.

(e) How do the classifier decision boundaries compare to the decision boundary you sketched as a result of visual inspection of data set 1? Explain why the boundaries produced by these three classifiers are similar, or different from, the decision boundary you sketched.