Introduction to Machine Learning

ECE 580
Spring 2022

HW #5

Submission Instructions

Submit your work to the corresponding assignment in Gradescope. Although Gradescope accepts multiple file formats, they strongly recommend submitting your assignment as a single PDF file.

You are responsible for ensuring the uploaded file is: 1) the correct file, 2) complete (includes all pages), 3) legible, and 4) submitted on-time as determined by the Gradescope server system clock.

You are responsible for tagging the pages that correspond to each question. Pages may be tagged after submission, even if the submission deadline has passed. If you are submitting close to the submission deadline, submit your assignment first then immediately return to tag the pages.

When code is requested, submit a PDF print-out of your code. Submitting a URL for a cloud-based repository is insufficient.

Scoring Information

Individual questions will be scored holistically on the 9-point scale described in the syllabus. The homework assignment score will be the weighted average of the individual scores for each question. (The weight for each question is shown in parentheses to the left of the question number.)

Late Submissions

Late submissions will be accepted up to 5 days after the submission deadline, with the following point penalty applied if its late submission is not excused: [1]

1 day (0^+ to 24 hours) late: 0.2 point deduction
2 days (24^+ to 48 hours) late: 0.5 point deduction
3 days late: 0.8 point deduction
4 days late: 1.6 point deduction
5 days late: 3.2 point deduction
6 or more days late: score = 3.0 (not accepted for credit)

The late policy is designed to be minimally punitive for submissions up to 3 days late, yet encourage staying current with the coursework for our course by not allowing one assignment's late submission to overlap with the next assignment's submission.

A homework score will not drop below 3.0 as a result of applying the late penalty point deduction.

[1] One day = one 24-hour period or fraction thereof.

Exploring Bayes Classifiers

Make sure you are able to apply a Bayes classifier and that you have flexibility to specify the assumptions you wish to make regarding the covariance structure of the data, *i.e.*, unique covariance matrices for each class, a common covariance matrix for both classes, independence among features (a diagonal covariance matrix), independence and common scaling among features ($\Sigma = \sigma^2\mathbf{I}$).

Regardless of whether you choose to write your own functions or leverage functions that may be available through standard Matlab or Python packages, it is critical that you understand 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.

(5) 1. Generate a 2-dimensional dataset for two distinct classes (class 0 and class 1) such that the dataset satisfies the following properties:

- the covariance matrix[2] for class 0 has a positive correlation, and $\sigma_0^2 \neq \sigma_1^2$
- the covariance matrix for class 1 has a negative correlation, and $\sigma_0^2 \neq \sigma_1^2$
- $\mu_0 \neq \mu_1$, but the distributions overlap noticeably

Provide the parameters for your dataset (μ_0, Σ_0, μ_1, Σ_1), and visualize it as a scatter plot of the data points, clearly indicating the associated class for each data point.

(5) 2. Apply a Bayes Classifier to your dataset, making no simplifying assumptions regarding the covariance structure (*i.e.*, the features may be dependent, and the covariance matrices are unique).

(a) Plot the decision statistic surface[3] with both the training data and the decision boundary under the assumptions of equal class priors and symmetric costs superimposed on top.

(b) How do the assumptions made regarding the covariance structure of the data while training the classifier affect the resulting decision boundary?

(5) 3. Apply a Bayes Classifier to your dataset, under the simplifying assumption that the features are independent but the covariance matrices under each class are unique.

(a) Plot the decision statistic surface with both the training data and the decision boundary under the assumptions of equal class priors and symmetric costs superimposed on top.

(b) How do the assumptions made regarding the covariance structure of the data while training the classifier affect the resulting decision boundary?

(5) 4. Apply a Bayes Classifier to your dataset, under the simplifying assumption that the covariance matrices under both classes are the same but the features may be dependent.

(a) Plot the decision statistic surface with both the training data and the decision boundary under the assumptions of equal class priors and symmetric costs superimposed on top.

(b) How do the assumptions made regarding the covariance structure of the data while training the classifier affect the resulting decision boundary?

[2]Recall that covariance matrices can be decomposed into components that characterize the "directions" in which the data lie and the strengths of those "directions", so you can define your covariance matrices in terms of this decomposition, and then synthesize the covariance matrices.

[3]Visualizing the ln-likelihood ratio is usually preferred as it is easier to interpret visually than the likelihood ratio.

(5) 5. Apply a Bayes Classifier to your dataset, under the simplifying assumptions that the covariance matrices under both classes are the same and the features are independent.

(a) Plot the decision statistic surface with both the training data and the decision boundary under the assumptions of equal class priors and symmetric costs superimposed on top.

(b) How do the assumptions made regarding the covariance structure of the data while training the 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)

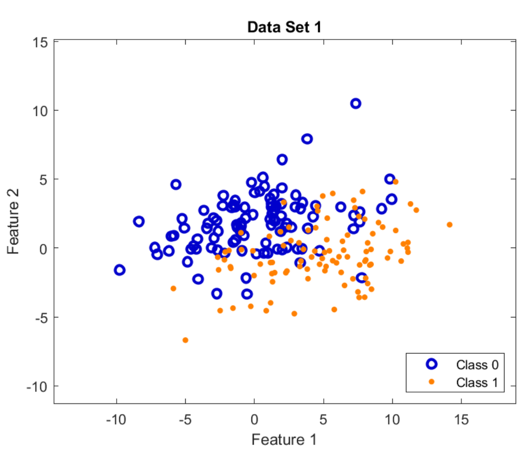
Linear Discriminant and Logistic Discriminant both assume a linear boundary separates the two classes; they differ in the assumptions they make to arrive at the resulting linear boundary and in how the coefficients for that boundary are computed. Here, you are going to explore how linear discriminant 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.

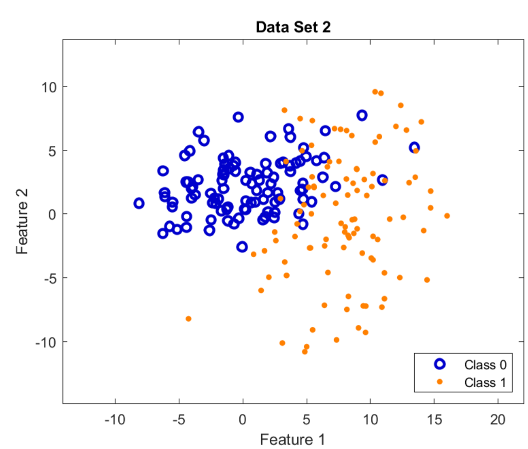
Make sure you are able to apply both a linear discriminant classifier and a logistic discriminant classifier.

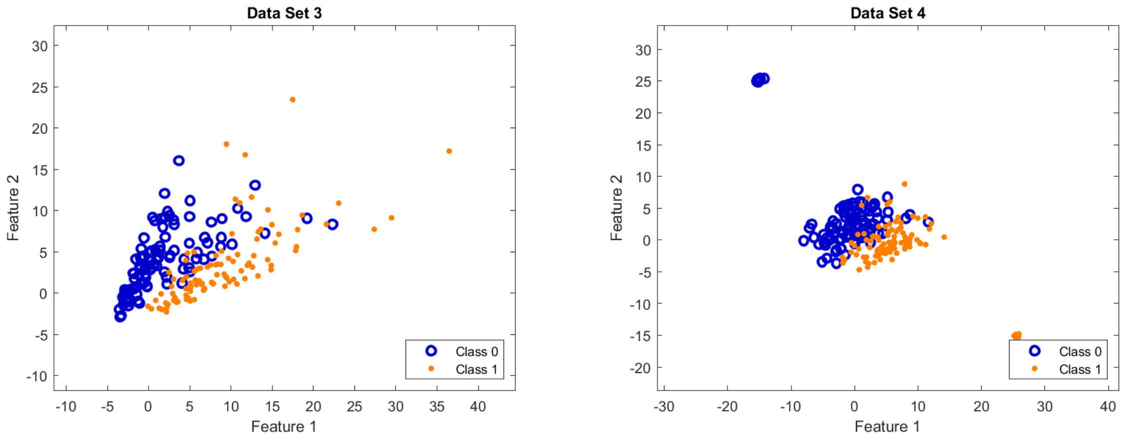
Regardless of whether you choose to write your own functions or leverage functions that may be available through Matlab or Python packages or libraries, you are responsible for understanding how the function(s) you are using work so you can effectively apply them to suit your needs and correctly interpret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

(10) 10. Data set 2 is not consistent with the assumptions underlying LDA – the data is Gaussian with means for the two classes that are distinct, but the covariances are also distinct.

(a) Apply the linear discriminant to dataset 2, 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 2, 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 2, 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 2 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 2? Explain why the boundaries produced by these three classifiers are similar, or different from, the decision boundary you sketched.

(10) 11. Data set 3 is not consistent with the assumptions underlying LDA – the data is not Gaussian, but rather is log-normal.

(a) Apply the linear discriminant to dataset 3, 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 3, 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 3, 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 3 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 3? Explain why the boundaries produced by these three classifiers are similar, or different from, the decision boundary you sketched.

(10) 12. Data set 4 is not consistent with the assumptions underlying LDA – the data is not Gaussian, but rather has outliers.

(a) Apply the linear discriminant to dataset 4, 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 4, 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 4, 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 4 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 4? Explain why the boundaries produced by these three classifiers are similar, or different from, the decision boundary you sketched.

(10) 13. Describe the factors you would consider when choosing among Bayes, Linear Discriminant, and Logistic Discriminant classifiers, and why you consider these factors important considerations.

(5) 14. Submit a PDF print-out of your code for this section (Comparing Linear Discriminant and Logistic Discriminant (and Bayes)). (Submitting a URL for a cloud-based repository is insufficient.)