

Homework # 2 (Generative Models) (100 points)

Due on March 12, 11:59pm EST

We will be working with the Penguin dataset again as we did for Homework #1. Please use “Species” as your target variable. For this assignment, you may want to drop/ignore the variable “year”.

Using the target variable, Species, please conduct:

- a. Linear Discriminant Analysis (30 points):
 - a. You want to evaluate all the ‘features’ or dependent variables and see what should be in your model. Please comment on your choices.
 - b. Just a suggestion: You might want to consider exploring featurePlot on the caret package. Basically, you look at each of the features/dependent variables and see how they are different based on species. Simply eye-balling this might give you an idea about which would be strong ‘classifiers’ (aka predictors).
 - c. Fit your LDA model using whatever predictor variables you deem appropriate. Feel free to split the data into training and test sets before fitting the model.
 - d. Look at the fit statistics/ accuracy rates.
- b. Quadratic Discriminant Analysis (30 points)
 - a. Same steps as above to consider
- c. Naïve Bayes (30 points)
 - a. Same steps as above to consider
- d. Comment on the models fits/strength/weakness/accuracy for all these three models that you worked with. (10 points)