

BMTRY Spring 2023: Assignment 3

Exercises

1. The data “LinSep” available on the class website are perfectly linearly separable.
 - i. Fit a linear regression and LDA model to these data. Show graphically that these models (1) are equivalent and (2) fail to select the “best” hyperplane to separate these data.
 - ii. Develop an R function to implement a basic perceptron algorithm. Arguments in the function should include predictors x , class variable y , initial choice of β 's, and learning rate.
 - a. Initialize the β 's to values of (0,0,0) and show that the perceptron outperforms the linear and LDA classifiers from part (i).
 - b. Show graphically that changing the initial choice of β 's impacts the location of the hyperplane selected by your algorithm.
2. An orthodontist at MUSC is interested in determining if information collected from 2-D CT scans is predictive of different types of root resorption in adults with periodontitis. The data includes scans from 60 unique canines and the primary outcome is root resorption defined in the categories: none, dentine involvement, or pulp involvement. Root resorption classification was ascertained using 3-D CT scans. Two dimensional CT scans were also conducted on each impacted canine and the following three measures were calculated from the 2-D scans: (1) U3-U2 angle, (2) the U3 tip-midline (mm), (3) the U3 tip horizontal reference (mm). Download the data from the class website.
 - i. Select a “best” linear regression, multinomial logistic regression, and LDA classifiers to predict root resorption class. To do this, consider and compare the following models
 - a. The three models containing only 1 main effect.
 - b. The three models containing 2 main effects.

c. The model containing all main effects.

To select the “best” model for each method, use a 10-fold cross-validation approach where the outcome is the average misclassification rate.

- ii. Fit the “best” model for each method using the full data set and plot the observed resorption class by the two predictors. Indicate on the plot which observations are misclassified and to which class they are mis-assigned. Additionally, for the linear and multinomial regression models, plot the decision boundaries between each class.
- iii. What model could you use if root resorption was ordinal (i.e. None < Dentin < Pulp)? What are the assumptions for your proposed model?