

DS-630: HW3
Assigned Feb. 13, 2018
Due Feb. 27, 2018

In this assignment you will use the `lda` algorithm from the *MASS* package, the `knn` algorithm from the *class* package, and the `glm` function for logistic regression in R to classify malignant vs. benign breast tumors. The dataset you will use can be found at the UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets.html>, and is labeled *Breast Cancer Wisconsin (Diagnostic)*.

This dataset contains the ID, diagnosis, and 30 real-valued input features. You are to split the dataset into two pieces: a training set consisting of the first 75% of the observations, and a test set consisting of the remaining observations.

Please use your best judgement to preprocess your data appropriately and employ *all* three algorithms `lda`, `knn`, and `glm` to predict the correct diagnosis for as many test instances as possible. Report your findings along with a written explanation of your process and results. Note, be sure to compare and contrast your results employing each algorithm.