MLDL

* Assignment review
  + Q1: 3-3.6
  + Q2: correct code
  + Q3: GPA predictor
* Classification
  + Algorithms
  + See whether it is one of the specified categories
  + Model
    - Import
    - Encode
    - Define x and y
    - Split the data
    - Fit model
  + Confusion matrix
    - See how the algorithm is performing
    - Accuracy (whole model) =
    - For a particular class
      * Precision =
      * Recall =
      * F1 =
  + Code review
    - Y is going to be the encoded values only
    - Logistic regression algorithm
    - Accuracy =
    - Precision of setosa = 13/13 = 1
* KNN Classifier
  + Same code format
* Unsupervised learning
  + Unlabeled data
  + Clustering
    - Find patterns or groups in the data
* Clustering
  + Group the data
  + Trial and error
  + If the label is present, then it is classification
  + K-means algorithm
    - Separates in k non-overlapping groups
* DB san method
  + Finds the points near and put in one cluster
* Assignment 2 discussion
  + Work doc
* Applications
  + Marketing
  + Biology
  + Libraries
  + Insurance
* IEEE paper
  + How to know which method to implement
* Elbow method
  + Find the best value of K
  + Inertia: sum of the squared distances of samples to their closest cluster center
  + Distortion: average of the squared distanced from the cluster’s centers
  + Select the “elbow” of the graph Inertia X K