**CME 4403 Hw3 - Due Date: March 26th 2019, 23:59**

*Please send your source codes and results (figures, replies etc.) in one text file (e.g., YourName.docx / YourName.pdf), then submit it to the Classroom.*

*Group homework is not allowed!*

1. You will code a **decision tree** (DT) that will be similar with the one given in the lab examples. The input data is provided in the text file “wheat\_types.txt”. The target feature is the “type” column. You will train and test the DT by using all features except “type”. Apply the following steps and reply the questions in your report.

* You can use a different decision tree library apart from “tree”.
* Use 80% of samples for training and 20% of them for test purposes.
* Perform this sample division (cross-validation) process for “100” iterations.
* Draw a plot to show how accuracy changes over 100 iterations.
* Report average accuracy of DT after 100 iterations.
* Show the DT (with parent and internal node’s decision criteria) which has the highest accuracy over 100 iterations.

1. You will code a **k-nearest neighbor** (kNN) algorithm, which will be similar with the one given in the lab examples. You will use built-in data set “Smarket” that is provided in the “ISLR” library. The target feature is the “Direction” column. You will use various combinations of descriptive features and different ***k*** values. Apply the following steps and reply the questions in your report.

* You need to test several combinations of descriptive features to improve the prediction accuracy of the target feature. The combinations:
  + Lag1, Lag2, Lag3
  + Lag3, Lag4, Lag5
  + Lag1, Lag2, Lag3, Lag4, Lag5
* Use 75% of samples for training and 25% of them for test purposes (for each combination).
* Test different ***k*** values (1 to 50) for each combination.
* Make a plot to show how accuracy changes for different ***k*** values (for each combination).
* Which ***k*** value did provide the highest accuracy for each combination? Report both ***k*** values and the highest accuracy value.
* What is the highest accuracy to predict the “Direction” column when you set a specific ***k*** value and the descriptive feature combination?