Problem Set 3

This problem set will be graded mainly on the precision rates from your machine learning models.

Q 1 Mobile Price Classification

In this question, you will be working on a dataset, **mobile_price.csv**. This data-set provide information between 20 features of a mobile phone (e.g., RAM, Internal Memory, Front Camera mega pixels, etc) and its selling price segments, from 0 to 3. Essentially, this is a classification problem with 4 classes. The class variable is **price_range**. Others are attributes. Please use various supervised learning algorithms to answer the following questions.

Use **test_size=0.30, random_state = 512** for **train_test_split** method in order to achieve consistent results.

a) Use KNN algorithm to train your machine learning model using train data-set generated from **train_test_split** method. Which k do you pick for your KNN model? Please provide a graph for k-value against error rates as we did in class. What is your highest precision rate for your KNN model? Please provide your classification report and confusion matrix. (Hint: For this question, you do NOT need to standardize the attributes.)

b) Use Decision Tree algorithm to train your model. What is your highest precision rate for your DT model? (Hint: You can play with different parameters to see if you can gain better predictions.)

c) Use Random Forest algorithm to train your model. What is your highest precision rate for your Random Forest model? (Hint: You can play with different parameters to see if you can gain better predictions.)

d) Use **LogisticRegression** in **sklearn.linear_model** and train your logit model with **LogisticRegression(multi_class='multinomial', solver ='newton-cg')**. This will provide you with multinomial logistic model which can classify multiple classes. What is your precision rate for your multinomial logistic model ?