**Süleyman Tuğrul Dinçer – 2095354**

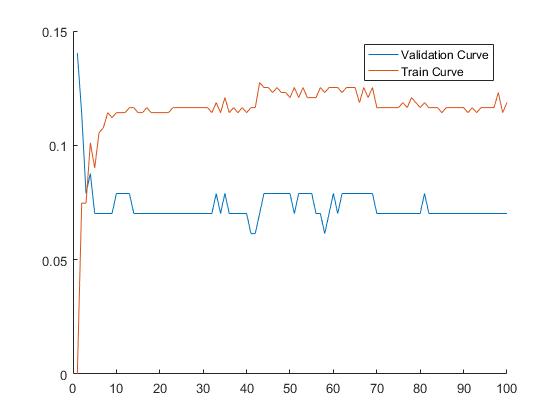
**KNN\_CODING\_ASSIGNMENT**

**PART\_I**

In this part of the coding assignment, we are given a dataset which is related to patients’ cancer situation with 10 different features, and class part for diagnosis. ‘1’ is used for diagnosing the disease, and ‘0’ is used for vice versa. Through the homework, I used the seed function as rng(3).

PART\_I

In this part of the assignment, we are given a build-in function as *fitcknn* of Matlab. Steps of the code is as follows:

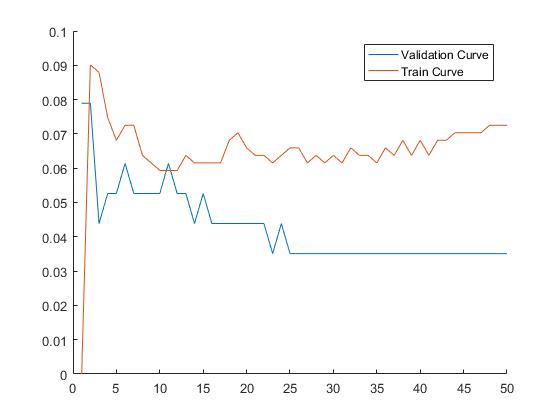
1. Dataset is shuffled by using randperm function, then it is separated into two groups as train and validation data. The partition is in the form of 80% for train set and 20% for validation set.
2. For the distance measure Euclidean distance is used.
3. There is no cross-validation for this assignment.
4. To predict validation classes, 100 different knn values are used from 1 to 100.
5. By using plot function we can observe how error differentiate values by changing knn values.
6. Optimal knn=41 and error rate is chosen as 0.0614.

**PART\_II**

In this part of the question, I came up with my own KNN function as *fuzzyknn* and *rnearest knn*.

a) I implemented fuzzy algorithm with Euclidean distance. I coded fuzzyknn function in another script, then called it to the main script to see the results of different knn value.

b) After classifying validation and train dataset, error rates are coming up with following plot. In that plot, Minimum error rate for validation data is 0.0351, and the optimal knn value for that error rate is 23.



c) When r-Radius Neighbours KNN Algorithm is implemented, optimal r value is chosen as 0.4, and the corresponding error rate is as 7%. For this purpose, I coded rnearest function in another script and called it to the main script to see the results of different knn values.

