Q1 (b) The theoretical time complexity of the KNN algorithm is O(mn)

→ # m is the number of samples &

→ # n is number of features

→ # My algorithm has a linear time complexity with respect to number of samples & number of feature. This agrees with the theoretical time complexity of KNN

Q2

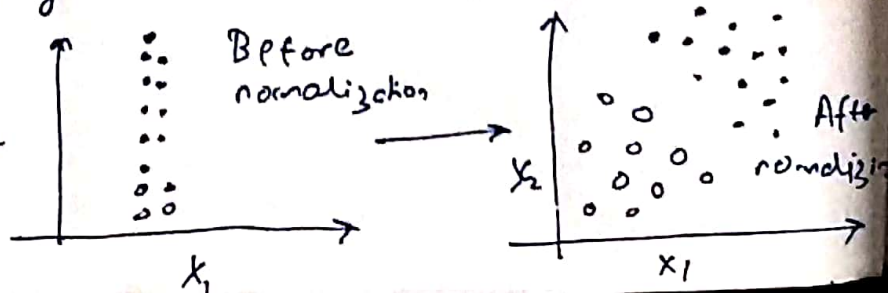
(a) The test error obtained using KNN gives slightly lower test errors compared to test errors in previous assignments.

(b)

Yes the scale of the input features impacts the performance of the KNN algorithm

This is because the features with significantly large variation of values would dominate the ones with less variation.

Normalizing the dataset reduces this variation.



Q2

(c) The Test error reduces & then increases. It has a parabolic shape.

The train error increases as K increases.

Q2 (c)

(i) Home number 102 M gives maximum error.

This is because only one parallax feature shows a significant change in the dataset.

This can be improved by normalizing the features between 0 & 1

Q3. The model leads to overfit the data thus 3 distinct classes are not distinctly separated.

This can be improved by increasing value of K.