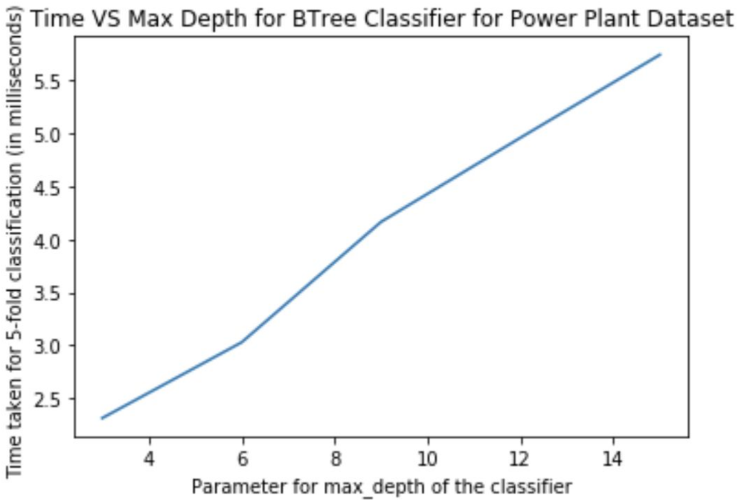
Question 1)
a) In a Decision Tree the induction algorithm at each node follows the greedy approach to split that particular portion of the data into subsets that are homogeneous as possible, hence the optimal ordering is contingent to the distribution of the data points.

b) MVE - 0.16089**, Best Parameter =** 9

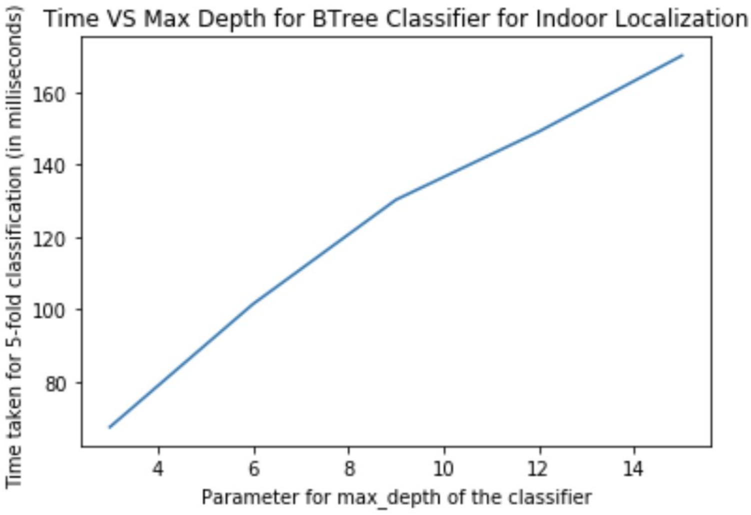


The classifier works comparably good on the test set as it reports a Mean Absolute Error of **0.16089** which is close to how it was performing in the Cross- Validation on the test set.

Parameter	**Mean Absolute Error**
3	**0.21208**
6	**0.17250**
9	**0.160034**
12	**0.162114**
15	**0.166038**

c) MVE - 11.16098 Best Parameter : 15

Parameter	**Mean Absolute Error**
3	**42.36554**
6	**25.69101**
9	**18.22498**
12	**12.36323**
15	**10.11299**

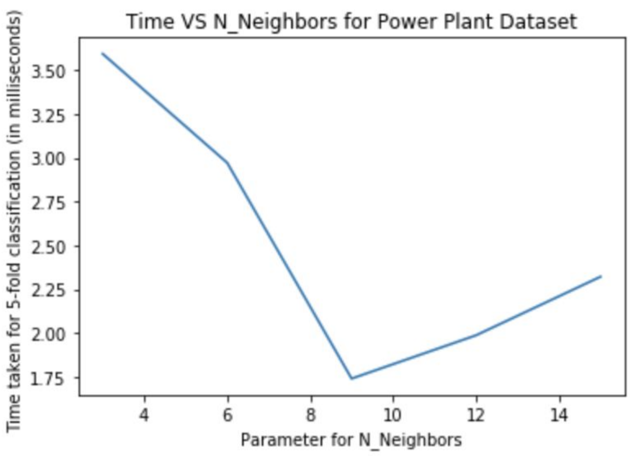


The classifier works comparably good on the test set as it reports a Mean Absolute Error of **11.16098** which is close to how it was performing in the Cross- Validation on the test set.

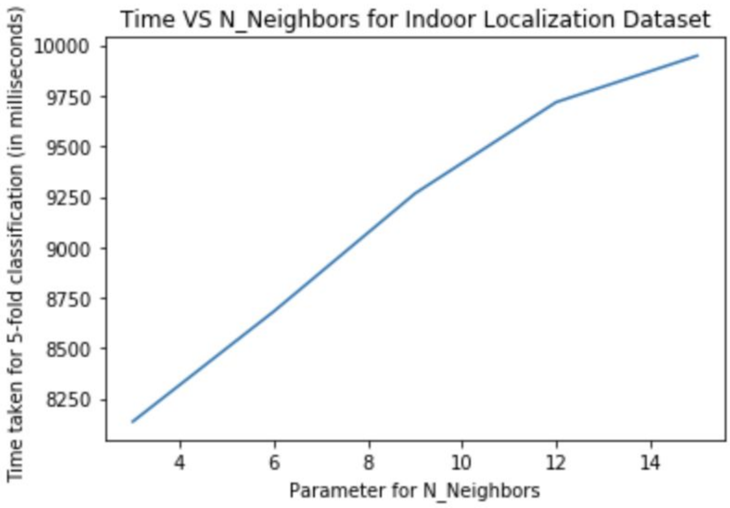
Question 2) KNN
a) Best Parameter : 3, **MAE:** 0.53095

Parameter	**Mean Absolute Error**
3	**0.541715**
5	**0.550275**
10	**0.579383**
20	**0.595314**
25	**0.598548**

The classifier works comparably good on the test set as it reports a Mean Absolute Error of **0.53095** which is close to how it was performing in the Cross- Validation on the test set.



b) Indoor Localization Best Parameter n_neighbor = 3 MAE: 3.16736



Parameter	Mean Absolute Error
3	**3.7628639851**
5	**4.1370138908**
10	**4.9786867405**
20	**5.9020269809**
25	**6.2183552425**

The classifier works good on the test set as it reports a Mean Absolute Error of **3.16736** which little better than how it was performing in the Cross- Validation on the test set.

Question 3

c) Larger (λ) means more regularization and w will be closer to 0 and towards each other
L2= Ridge. As the value of λ increases the model complexity decreases. This is more profound in L1 as compared to L2 the contributions of features fall off faster in L1 as compared to L2 as λ increases

Question 4 Lasso and Ridge Regression

a) We can get large values for the weights if we have a large number of parameters or when the columns are correlated. Regularization prevents overfitting by restricting the model and reducing the models complexity. The regularization reduces the unstable high estimates of w to be closer to zero, introducing bias but reducing the variance of the estimates.

b) Ridge: Best Alpha = 10e-4 **MAE :** 0.19137

Parameter (alpha)	Mean Absolute Error
10e-6	**0.190864851**

3a) Given Equation

$$G(w,\lambda) = (w-2)^2 + (w-4)^2 + \frac{\lambda}{2} w^2$$

$$\frac{\partial G}{\partial w} = 2(w-2) + 2(w-4) + \lambda w$$

$$= 4w - 12 + \lambda w$$

$$= (4+\lambda) w - 12$$

Critical points $(4+\lambda)w - 12 = 0$

$(4+\lambda) w = 12$

$$\boxed{w = \frac{12}{4+\lambda}, \lambda > -4}$$

$$\frac{\partial^2 G}{\partial w^2} = (4+\lambda)$$

for G to be minimizes $\frac{\partial^2 G}{\partial w^2} > 0$

$(4+\lambda) > 0$ if $\therefore$ $\lambda > -4$

b) Given $H(w,\lambda) = (w-2)^2 + (w-4)^2 + \lambda|w|$

Case I $w \geq 0$

$$\frac{\partial H}{\partial w} = 2(w-2) + 2(w-4) + \lambda$$

$$= 4w - 12 + \lambda$$

Critical point $w = \frac{12-\lambda}{4}$

$\frac{\partial^2 H}{\partial w^2} = 4 > 0$ hence $\frac{12-\lambda}{4}$ is minimum for all values of λ

10e-4	**0.190864852**
10e-2	**0.190864866**
1	**0.19086500**
10	**0.1908663594**

The classifier works comparably good on the test set as it reports a Mean Absolute Error of **0.19137** which is close to how it was performing in the Cross- Validation on the test set.

Lasso:
Best Alpha = 10e-4 **MAE :** 0.19137

Parameter (alpha)	**Mean Absolute Error**
10e-6	**0.1908649**
10e-4	**0.1908720**
10e-2	**0.194185**
1	**0.2627**
10	**0.627926**

The classifier works comparably good on the test set as it reports a Mean Absolute Error of **0.19137** which is close to how it was performing in the Cross- Validation on the test set.

Best Model was Ridge : with an alpha of **10e-4** but on further analysis both Lasso and Ridge 10e-4 gave the same Mean Absolute Error.

(Indoor Localization)
Ridge: Best Alpha = 10 **MAE :**20.69706

Parameter (alpha)	**Mean Absolute Error**
10e-6	**19.65125**
10e-4	**19.64871**
10e-2	**19.61162**
1	**19.60135**
10	**19.58408**

The classifier works comparably not as good on the test set as it reports a Mean Absolute Error of **20.69706** which is a little off from the error on the test set validation.

Lasso:
Best Alpha = 0.001 **MAE :** 20.67810

Parameter (alpha)	**Mean Absolute Error**

10e-6	**19.60657**
10e-4	**19.59844**
10e-2	**19.72880**
1	**21.45238**
10	**35.5881**

The classifier works comparably not as good on the test set as it reports a Mean Absolute Error of **20.67810** which is a little off from the error on the test set validation.

Question 5 Kaggle

a) Best = DecisionTreeRegressor (max_depth=12, min_samples_leaf= 5)
CV=5
MAE = 0.15877

Other Parameters tried:
DecisionTreeRegressor (max_depth=15, min_samples_leaf= 5) cv= 5, 8
DecisionTreeRegressor (max_depth=[1,3,4,5,6,7,8,9], min_samples_leaf= 5) cv= 5,8
DecisionTreeRegressor (max_depth=[1,3,4,5,6,7,8,9], min_impurity_decrease= [1,2,0.1,0.11]) cv= 5,8

b) Best = KNeighborsRegressor(n_neighbors = 1, n_jobs=-1, weights= 'distance', p=1)
Cross Validation = 5
MAE= 2.43565

Other parameters I have tried : KNN n_neighbors = 3,4,5,6,8,9,10 weight = ‘uniform’ p=2 in varying combinations.