VIGNESH PRABHAKAR
USC ID: 4024890667

9/24/19

Homework (2):

2.4.1

(a) We would expect a flexible statistical learning method to be better when sample size is extremely large & no. of predictors is small. this is because when we are provided with lower dimensional data but with lot of data point instances in those lower dimensions; a non flexible method models the relationship between the predictors & response poorly and would underfit our data due to a very high availability of data points in a smaller space. On the contrary a flexible statistical learning method would adjust it's wigglyness in order to accomodate majority of the data points in the model to yield an ideal fit.

(b) We would expect a non flexible statistical learning method to perform better when sample size is small but the no. of predictors is large. this is due to the curse of dimensionality. In order to provide an accurate estimate $\hat{f}(x)$ of the function that describes the relationship between our predictors & response in higher dimensional space; we would need extremely large sample size. Since that's

not the case ; we assume that estimating the parameter/ coefficients of the non flexible model would closely mimick the estimation of $\hat{f}(x)$ which is the conditional expectation of Y given X $\{E(Y|X=x)\}$. Due to the same reason of this assumption; a non flexible learning method is expected to perform better.

(c) Flexible statistical learning is expected to perform better when relationship between predictors & response is highly non linear. This is because we seek to obtain an ideal fit for majority of our data points. A flexible statistical learning method with less wigglyness would yield an ideal fit whereas a non flexible statistical learning method would underfit our data

(d) Non flexible statistical learning is expected to perform better when variance of error terms is high. High variance of error terms leads to overfitting thereby a flexible statistical learning method would perform poorly on generalizations. So; a non flexible statistical learning method is preferred in order to accomodate these generalizations as well and avoid overfitting.

2.4.7

(a) Let test data point $(0,0,0)$ be P.

then :

$E(P, i) \Rightarrow$ { euclidian distance between P & i^{th} observations }

$$E(P,1) \Rightarrow \sqrt{(X-x_1)^2+(Y-x_2)^2+(Z-x_3)^2}$$ { where (X,Y,Z) represent $(0,0,0)$ }

$$\Rightarrow \sqrt{(-3)^2}$$

$\Rightarrow$ 3 units

$E(P,2) \Rightarrow$ 2 units

$E(P,3) \Rightarrow \sqrt{1+9} = \sqrt{10} = 3.162$ units

$E(P,4) \Rightarrow \sqrt{1+4} = \sqrt{5} = 2.236$ units

$E(P,5) \Rightarrow \sqrt{1+1} = \sqrt{2} = 1.414$ units

$E(P,6) = \sqrt{1+1+1} = \sqrt{3} = 1.732$ units

(b) with $k=1$,

the nearest neighbor to our test data point is 5^{th} observation. { 1.414 units }

therefore the class of 5^{th} observation is assigned to the test data point which is green.

This is because the only neighbor where rolling matters is the most nearest neighbor; so we directly assign it's class to the Test data point Although this would be a bad approach due to the high variance that leads to overfitting of our model at lower values of K

(c) The 3 nearest neighbors with $k = 3$ are:

* observation 5 => 1.414 units
* observation 6 => 1.732 units
* observation 2 => 2 units

Observation 2 and observation 6 belong to red class whereas observation 5 belongs to green class.

By majority polling {default assumption unless specified otherwise} we get:

red to be the predominant class.

Therefore we assign red class to our test data point P(0,0,0).

d) The Bayes decision boundary closely mirrors our KNN decision boundary with similar test error ~~rates~~ as our KNN decision boundary at a given value of K. We know that Bayesian decision boundary

produced by a bayes classifier yields the lowest possible test error rate as we always choose the class for which $Pr(y=j \mid X=x_0)$ is maximum. We have also been given the information that such a bayesian decision boundary is "highly non linear". This situation arises at smaller values of K in a KNN model. The lower the value of K; the more non linear is the decision boundary and the higher the value of K; the less non linear is the decision boundary. So; a highly non linear decision boundary (bayesian) that yields the best test error rate; can happen at smaller values of K.