ANISH RAJESH ADNANI USC ID: 4092610491
USC Email : adnani@usc.edu
github : midnightbot

ISLR

2.4

1. For each parts (a) through (d) indicate whether we would generally expect the performance of a flexible statistical learning ~~model~~ method to be better or worse than an flexible method. Justify your answer.

(a) The sample size n is extremely large, and the number of predictors p is small.

→ Flexible Method performance will be better in such a case.
Increasing the sample size will increase the variance, if variance is increased the prediction model should be more wiggly to accomodate hence flexible model should be used to hammer the curve more to accomodate high variance

(b) The number of predictors p is extremely large, and the number of observations 'n' is small?

→ Inflexible model performance will be better in such a case.

With small dataset and large predictors, there is high probability that flexible model will overfit as it will try to hammer the curve to fit to all the datapoints with 'n' being small there is a high chance flexible model becomes like a lookup table, hence failing to generalize the data.

(c) The relationship between predictors and response is highly non-linear?

→ Flexible model performance will be better in such a case.

As we know relationship between predictors and response is highly non-linear flexible model will perform well as non-linear relationship will include wiggliness in the data and hence model with higher degree terms is needed to accomadate such behaviour.

(d) The variance of the error term $\sigma^2 = Var(\epsilon)$ is extremely high.

→ Inflexible model performance will be better in such a case.

variance of the error term σ^2 is extremely high which indicates there are a lot of noise / outliers. Hence using a flexible model, will lead to extreme hammering of the curve, which will fail to generalize the overall idea.

ISLR

2.7 The table below provides a training data set containing six observations, three predictors, and one qualitative response variable.

Obs	X_1	X_2	X_3	Y
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	green
5	-1	0	1	green
6	1	1	1	Red.

Suppose we wish to use this data set to make a prediction for Y when $X_1 = X_2 = X_3 = 0$ using K-nearest neighbors.

(a) Compute the Euclidean distance between each observation and test point $X_1 = 0$ $X_2 = 0$ $X_3 = 0$

→ Euclidean dist = $\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + \ldots}$

Obs	dist
1	$\sqrt{(0-0)^2 + (0-3)^2 + (0-0)^2} = 3$
2	$\sqrt{(2-0)^2 + (0-0)^2 + (0-0)^2} = 2$
3	$\sqrt{(0-0)^2 + (1-0)^2 + (3-0)^2} = 3.1622$
4	$\sqrt{(0-0)^2 + (1-0)^2 + (2-0)^2} = 2.2360$
5	$\sqrt{(-1-0)^2 + (0-0)^2 + (1-0)^2} = 1.414$
6	$\sqrt{(1-0)^2 + (1-0)^2 + (1-0)^2} = 1.732$

(b) what is our prediction with $k=1$? why?

$$\rightarrow \text{MIN}(\underset{①}{3}, \underset{②}{2}, \underset{③}{3.1622}, \underset{④}{2.2360}, \underset{⑤}{1.414}, \underset{⑥}{1.732})$$

with $k=1$
the given point is nearest to test data point ⑤ which is $(-1,0,1)$, which is green hence with $k=1$

given point $(0,0,0)$ will also be assigned green

(c) what is our prediction of $k=3$? why?

$\rightarrow$ dist $= [\underset{①}{3}, \underset{②}{2}, \underset{③}{3.1622}, \underset{④}{2.2360}, \underset{⑤}{1.414}, \underset{⑥}{1.732}]$

sorting dist $= [\underset{⑤}{1.414}, \underset{⑥}{1.732}, \underset{②}{2}, \underset{④}{2.2360}, \underset{①}{3}, \underset{③}{3.1622}]$

nearest 3 point ares

⑤ $(-1,0,1)$ green
⑥ $(1,1,1)$ Red
② $(2,0,0)$ Red.

Since majority of these 3 labels is Red
Red is ~~to~~ assigned to given data point $(0,0,0)$

(d) If the Bayes decision boundary in this problem is highly non-linear, then would we expect the best value of K to be large or small? why?

→ The best value of K will be small.
Bayes decision boundary is highly non-linear which indicates there is high variance in data, which means we need a wiggly fit.
For a wiggly fit we need a smaller value of K to make the curve more wiggly rather than linear as in large value of K.