SIDDHANT NADKARNI ID: 2596-6618-89

HW#2 (INF-552)

2.4.1

b) The performance would be better if n is large and p is small because it would fit the data more accurately and result in a better fit

b) When p is large and n is small the flexible learning model would overfit the data and thus would a worse model than inflexible

c) flexible learning method would fit the data more accurately because of more degrees of freedom and thus would be a better fit

d) Flexible methods would be worse if $\text{var}(\epsilon)$ is large because it would overfit and fit the noise in this case instead of generalizing and being robust to noise

2.4.7

a)

Obs	X_1	X_2	X_3
1	0	3	0
2	2	0	0
3	0	1	3
4	0	1	2
5	-1	0	1
6	1	1	1

Euclidean distance

$\sqrt{3^2} = 3$

$\sqrt{2^2} = 2$

$\sqrt{10} = 3.2$

$\sqrt{5} = 2.2$

$\sqrt{2} = 1.41$

$\sqrt{3} = 1.7$

b) for k=1 Obs # 5, Y = [green] would be the prediction because it is the closest neighbour to it

c) if k=3, Obs # 5, #6, #2 would be the closest. Obs #2 and #5 are Red and #6 is green. Therefore prediction would be [Red] with prediction because it appears in majority

d) If Bayes decision boundary is non-linear, then k=[small] would present a more non-linear boundary than k=large. Thus best value for k would be [small]