2

ISLR 2.4.1

a) n extremely large p small

n very large

flexible model will, with less predictors, fit the large data set better

Inflexible method, riged, might not fit the dataset as good as flexible approach

flexible better

b) p very large n small

as n small, flexible model will overfit & perform badly for test data

inflexible model won't overfit & will better predict test data

flexible not good here

inflexible better

c) Nonlinear → highly

flexible model will attempt to fit this highly non linear data better than an inflexible model

flexible better than inflexible here

---

d) $\sigma^2 = Var(\varepsilon)$ is very high

flexible models try to fit this error term variance

This makes them worse as they will increase overall variance

flexible bad

As data shows a lot of variation that can't be modelled, flexible model will attempt to fit it to no avail, reducing overall model accuracy.

3 ISLR 2.4.7

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

To predict Y by KNN when $X = (0, 0, 0)$

a) $d((0,3,0), (0,0,0)) = 3$

$d((2,0,0), (0,0,0)) = 2$

$d((0,1,3), (0,0,0)) = \sqrt{10} = 3.16227$

$d((0,1,2), (0,0,0)) = \sqrt{5} = 2.23606$

$d((-1,0,1), (0,0,0)) = \sqrt{2} = 1.41421$

$d((1,1,1), (0,0,0)) = \sqrt{3} = 1.73205$

b) Prediction with $K = 1$

Green (It is closest to $(0,0,0)$).

c) $K = 3$

Red

Pts 2, 5, 6 closest

2 → red
5 → Green
6 → red

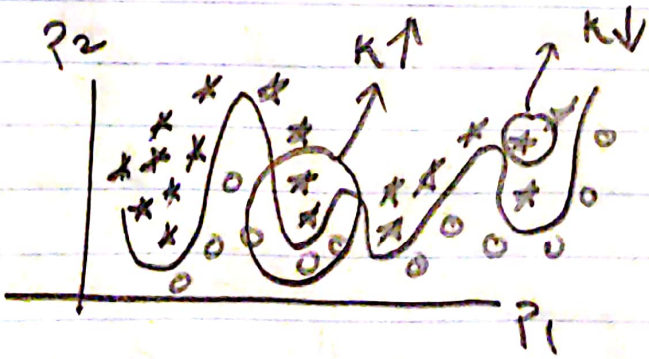
} majority polling → Red

d) Small K value better

Why ?

Non-linear nature high ↑
data set is behaving "locally"

Very linear then large K helps as data split uniformly "globally"



From fig.

K ↑ error ↑

K ↓ error ↓ (localized)

So K small → nonlinear highly
K large → if linear