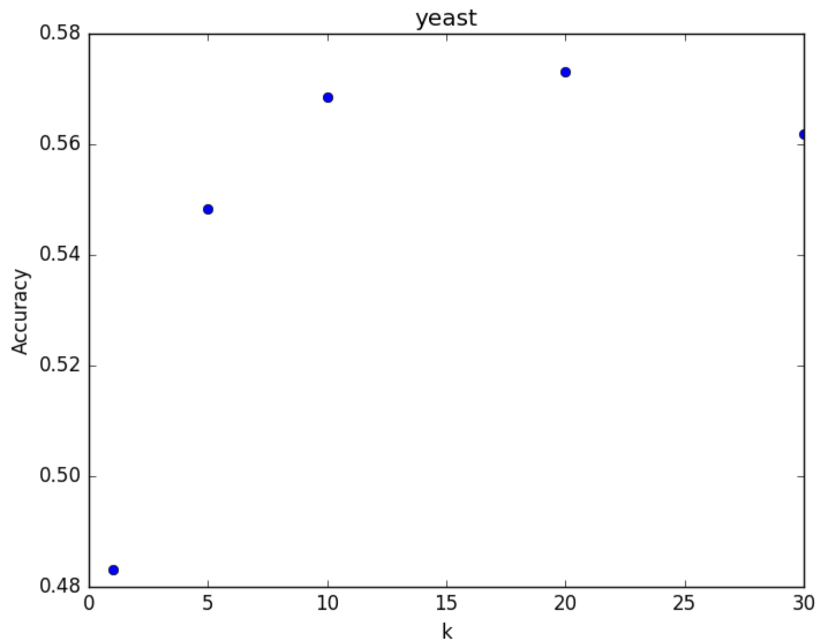
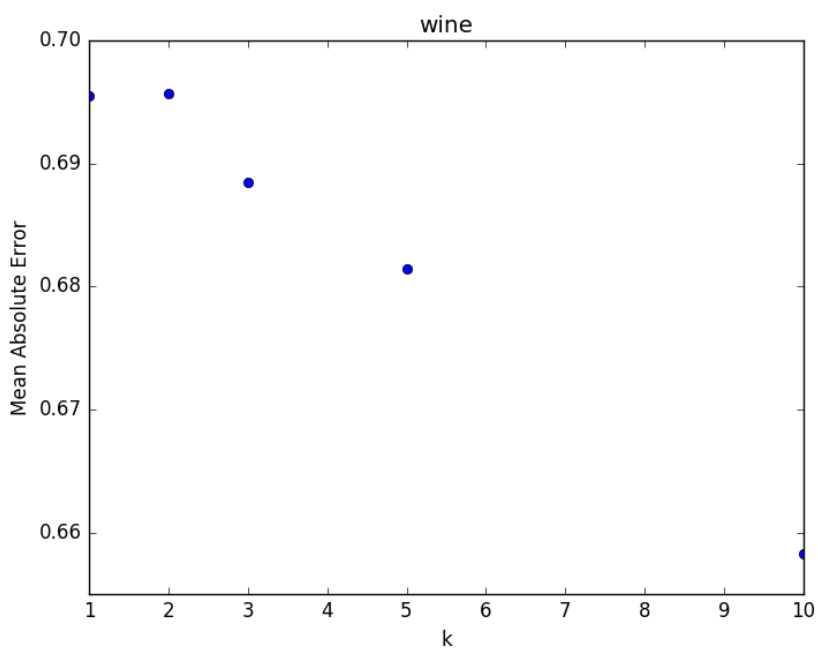
Part 2

1



yeast
0.58
0.56
0.54
0.52
0.50
0.48
Accuracy
0
5
10
15
20
25
30
k

2



wine
0.70
0.69
0.68
0.67
0.66
Mean Absolute Error
1
2
3
4
5
6
7
8
9
10
k

3

$k = 1$:

Predicted \ Actual	CYT	NUC	MIT	ME3	ME2	ME1	EXC	VAC	POX	ERL
CYT	69	46	26	3	3	0	0	3	1	0
NUC	39	57	12	9	0	0	0	0	1	0
MIT	21	13	34	2	2	1	1	1	1	0
ME3	4	6	2	32	3	0	0	0	0	0
ME2	1	0	3	1	6	3	1	0	0	0
ME1	0	0	0	0	1	6	3	1	0	0
EXC	1	0	2	0	2	3	5	0	0	0
VAC	2	1	0	0	0	0	0	1	0	0
POX	1	0	2	0	1	1	0	0	5	0
ERL	0	0	0	0	0	0	0	0	0	0

$k = 30$:

Predicted \ Actual	CYT	NUC	MIT	ME3	ME2	ME1	EXC	VAC	POX	ERL
CYT	89	48	25	3	4	0	0	2	6	0
NUC	36	59	5	2	2	0	0	3	0	0
MIT	11	13	45	2	3	1	1	1	2	0
ME3	1	3	1	40	3	0	0	0	0	0
ME2	0	0	1	0	3	1	0	0	0	0
ME1	1	0	3	0	3	9	4	0	0	0
EXC	0	0	1	0	0	3	5	0	0	0
VAC	0	0	0	0	0	0	0	0	0	0
POX	0	0	0	0	0	0	0	0	0	0
ERL	0	0	0	0	0	0	0	0	0	0

Observations:

- The correctly classified number of instances for the classes CYT, NUC, MIT, ME3, ME1 *increases* as k increases.

  We have a large number of instances for these classes, so larger k tends to give a more accurate classification.

- The correctly classified number of instances for the classes ME2, EXC, VAC, POX, ERL *does not increase* as k increases.

  We only have a very limited number of instances for these classes, so larger k does not necessarily yield better result due to the overfitting issue.

Part 3

pop	dist	best dist	best node	PQ
-	∞	∞	-	(f, 0)
f	7.0711	7.0711	f	(h, 0); (c, 1)
h	7.0711	7.0711	f	(i, 0); (c, 1); (g, 5)
i	3	3	i	(c, 1); (j, 3); (g, 5)
c	2	2	c	(e, 0); (b, 0); (j, 3); (g, 5)
e	7.8102	2	c	(b, 0); (d, 0); (j, 3); (g, 5)
b	4.4721	2	c	(d, 0); (j, 3); (a, 4); (g, 5)
d	5.3852	2	c	(j, 3); (a, 4); (g, 5)
j	END WHILE			

RETURNS c.