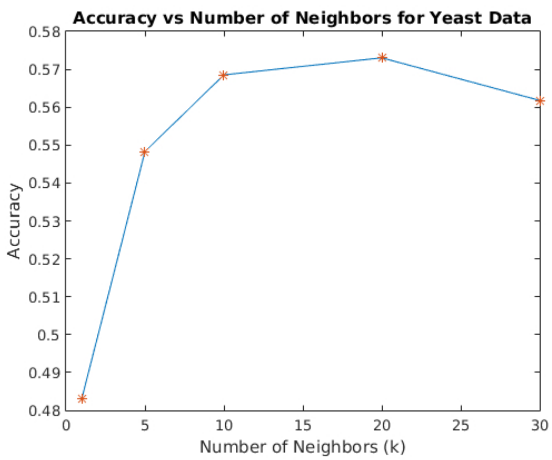
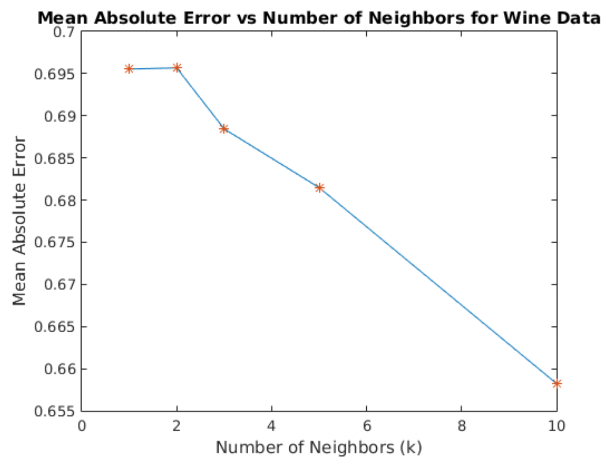
Colin Samplawski
CS 760 - Homework 2

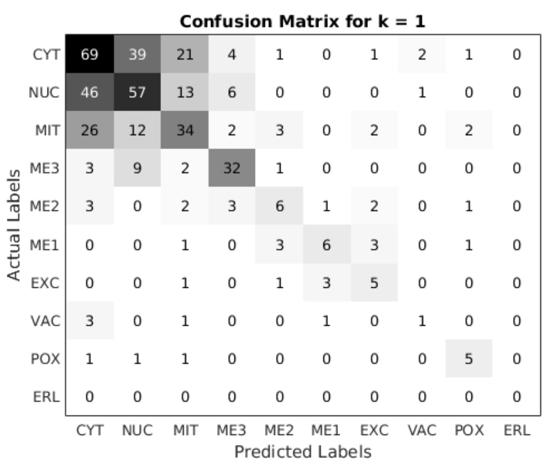
Part 1

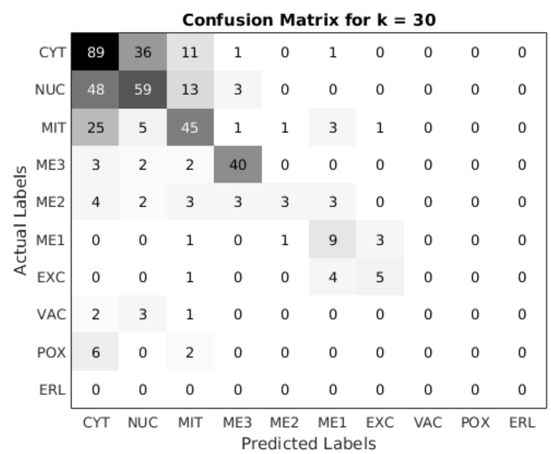
See included code.

Part 2









It easy to see that with using $k = 30$, there is less misclassified instances overall. This is not surprising given the graph above. It seems that most of the gains in accuracy where found in the better classification of instances with the CYT label, which is one of the more common labels in this data set. In fact, some other labels saw worse classification, such as NUC and MIT, but the gains in CYT compensated for it. Another interesting aspect of this data is that no testing instances had the ERL label and no instance was classified as such.

Part 3

operation	dist at this step	best dist	best node	queue
init	N/A	∞	N/A	$(f, 0)$
pop f	7.071	7.071	f	$(h, 0), (c, 1)$
pop h	7.071	7.071	f	$(i, 0), (c, 1), (g, 5)$
pop i	3	3	i	$(c, 1), (j, 3), (g, 5)$
pop c	2	2	c	$(b, 0), (e, 0), (j, 3), (g, 5)$
pop b	4.472	2	c	$(e, 0), (j, 3), (a, 4), (g, 5)$
pop e	7.810	2	c	$(d, 0), (j, 3), (a, 4), (g, 5)$
pop d	5.385	2	c	$(j, 3), (a, 4), (g, 5)$
pop j	Return c			