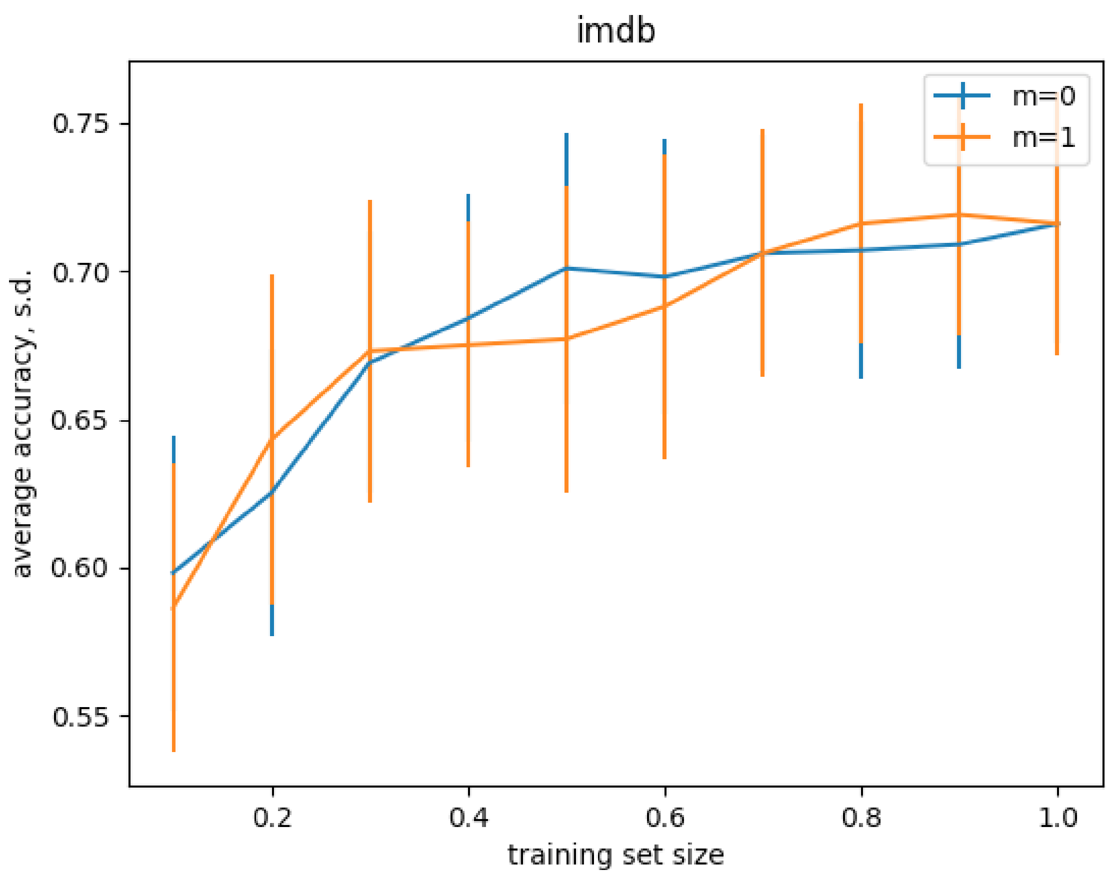
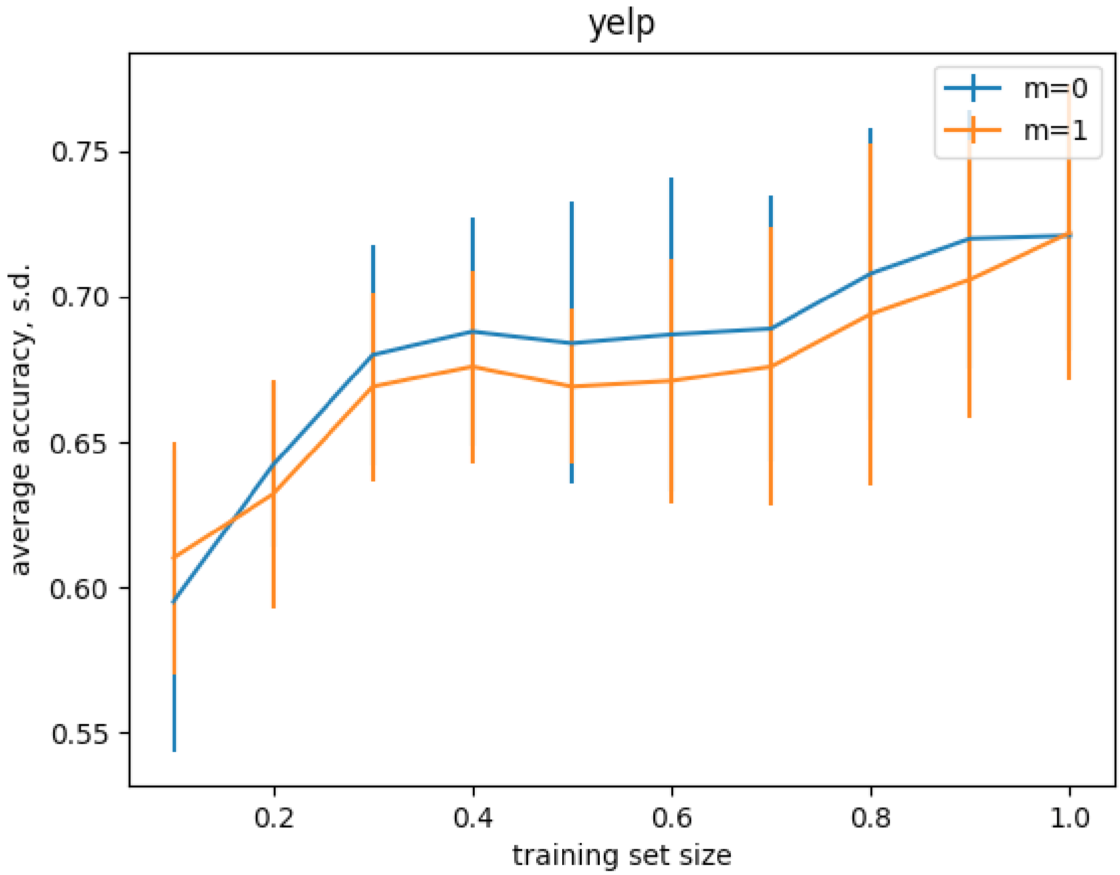
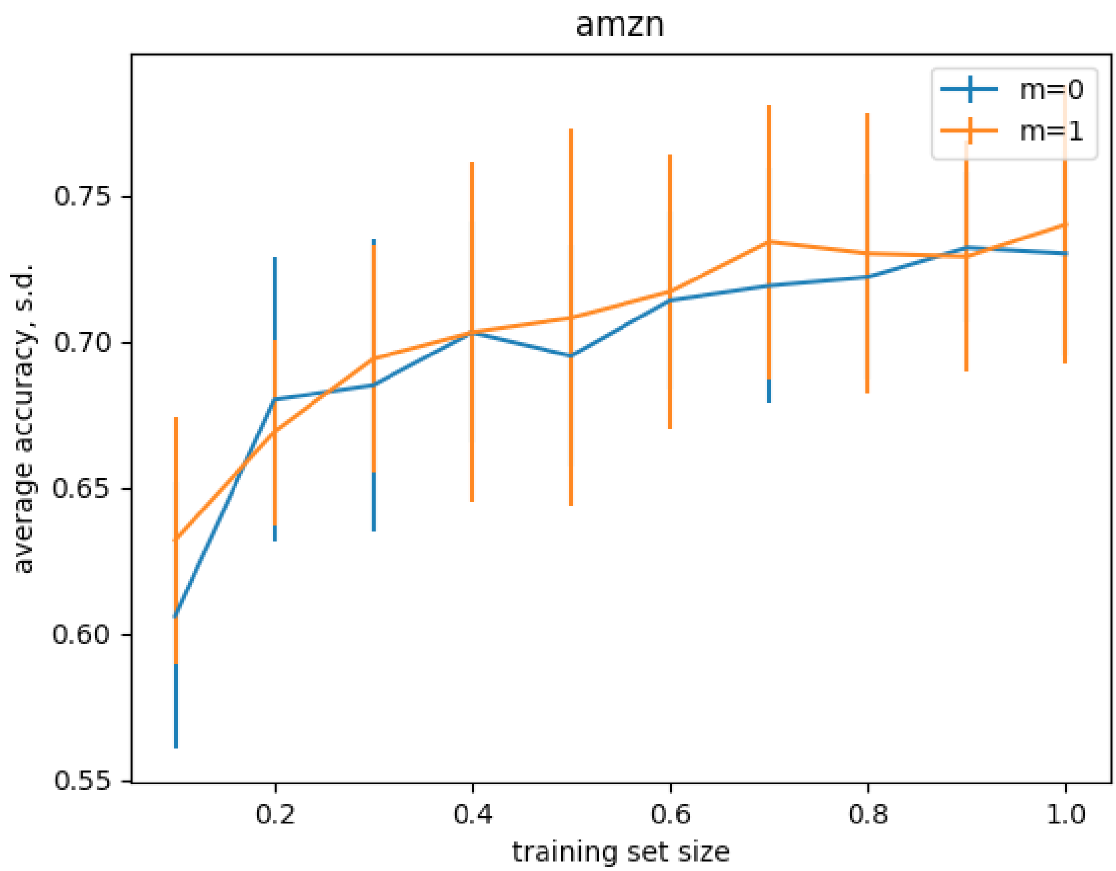
PP1 Report
Brendan McShane

Part 1

For part 1 we were testing how different training set sizes would affect accuracy and it's variance. Looking at all 3 resulting graphs (imdb, yelp, amzn), we can tell that the more data we're allowed to train on, the more accurate and reliable we were (as in we averaged a lower standard deviation across our k folds). This tends to be a trend with most machine learning models. The more data we have, the better our models tend to be to a point. Our graphs end up plateauing a little towards the end of the graph where we're using the full dataset, which implies that each individual extra data point gets less and less useful the more data we already have. Both maximum likelihood and the MAP approach performed similarly, although it would appear that MAP performs a little bit better with limited data. This makes sense with respect to both what we've discussed in class (maxL's tendency to overfit on insufficient data) and the smoothing process mentioned in the homework pdf, which helps us avoid 0/1 extreme solutions.



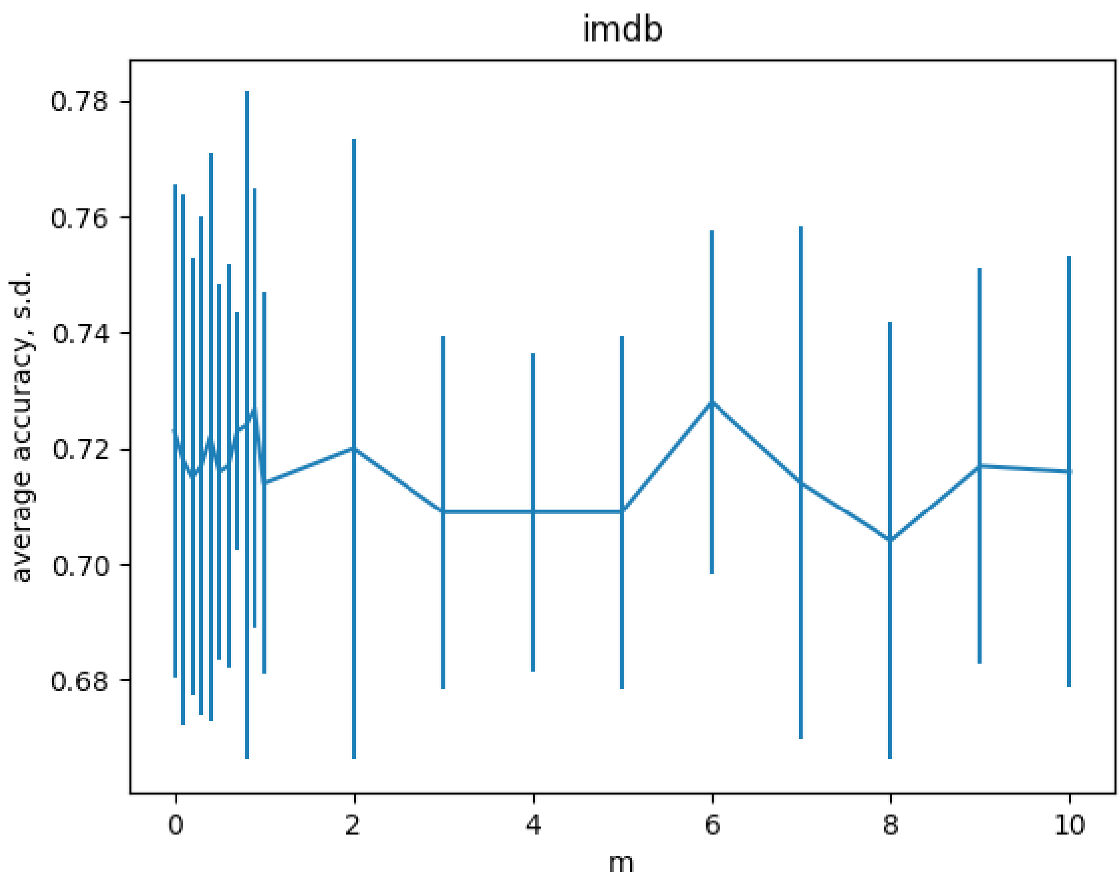

yelp
m=0
m=1
0.75
0.70
0.65
0.60
0.55
average accuracy, s.d.
0.2
0.4
0.6
0.8
1.0
training set size

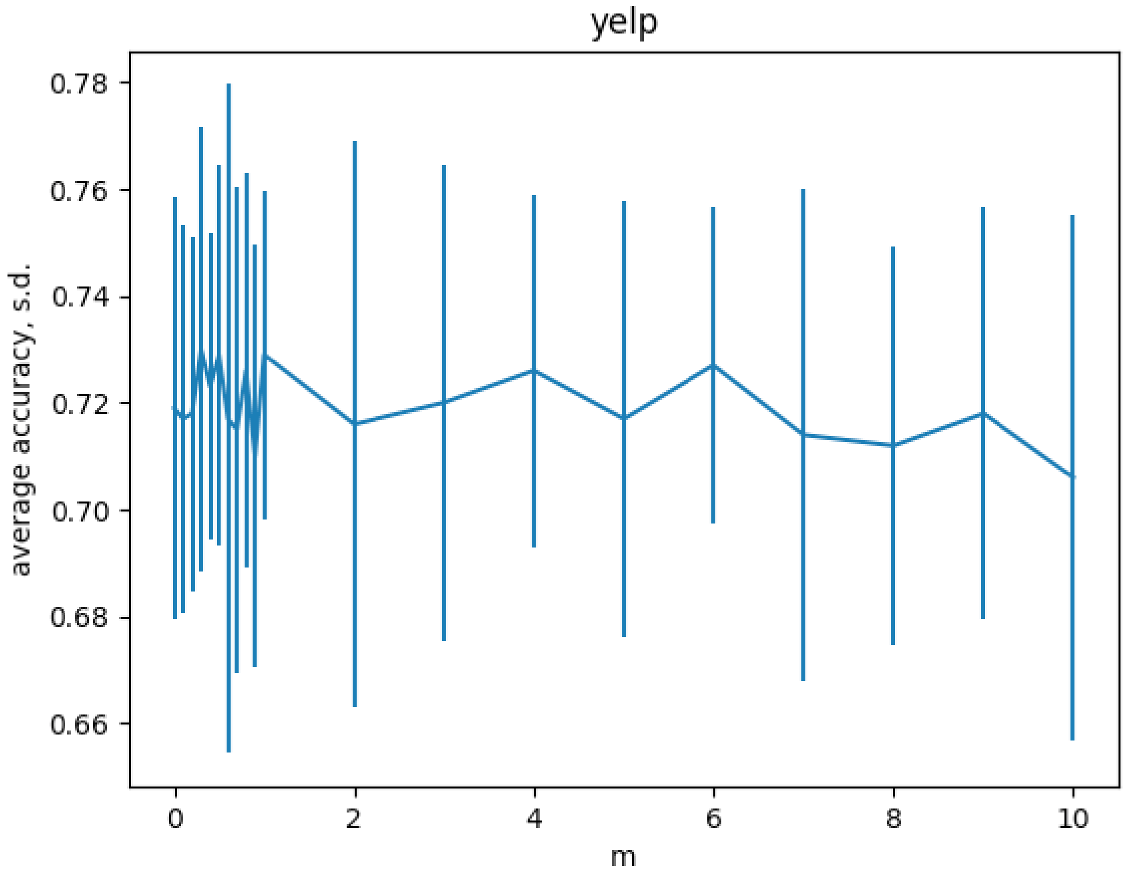


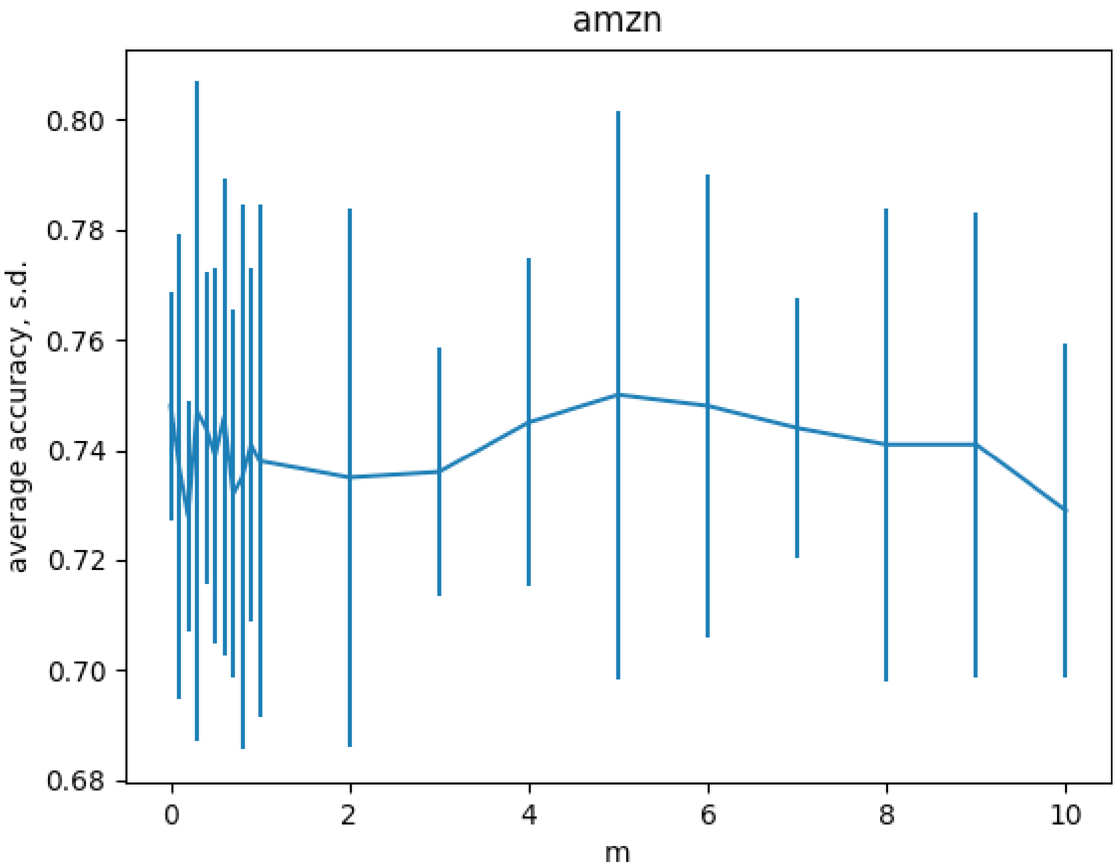
amzn
m=0
m=1
0.75
0.70
0.65
0.60
0.55
average accuracy, s.d.
0.2
0.4
0.6
0.8
1.0
training set size

Part 2

To be frank, I'm not sure what to make of these graphs. It would appear that the smaller the m value is, the more volatile our results are (notice how much the accuracies and s.d.'s are jumping all over the place in the m = [0,1] range). This makes sense to me as the entire point of the m-value (which behaves much like pseudocounts) is to smooth out our maximum likelihood values and avoid 0/1 extreme solutions. The accuracy past the m = [0,1] range seems to peak about halfway in the m = [4,6] range, and then drop off. The reliability of our results also seems to be at its best in the 4-6 range as our s.d.'s appear to be smaller in the middle of the graph. This implies relying too heavily on a large m value can actually hinder our results instead of help.




yelp
0.78
0.76
0.74
0.72
0.70
0.68
0.66
average accuracy, s.d.
0
2
4
6
8
10
m


amzn
average accuracy, s.d.
0.80
0.78
0.76
0.74
0.72
0.70
0.68
0
2
4
6
8
10
m