

570 HW10

Daniel Campos dacampos@uw.edu

12/06/2018

1 Q2

Expt	rare	feat	train accuracy	test accuracy	of feats	of kept feats	running time
1 ₁	1	1	95.92	82.60	39380	39380	4:42
1 ₃	1	3	96.99	83.42	39380	17703	5:47
2 ₃	2	3	97.70	84.65	41151	14974	6:37
3 ₅	3	5	97.19	83.54	40798	9511	6:57
5 ₁₀	5	0	97.30	79.87	40322	5100	6:56

Looking at table there are a few conclusions I can draw both about my program's performance and about the effect of the rare and feature threshold. As we increase the feature size we slowly grow the model size and increase the model performance. This is likely because the model can learn the true distribution better. As we increase the rare threshold we start dropping a larger portion of our features which eventually causes our model not to generalize as well. My program is not super efficient because in 1 1, the feature generation portion takes less than 5 seconds and by the time we get to 5 10 it takes 4 minutes. The model has less features to learn so it can train quicker but the generation of the vectors takes longer. Given more time I would optimize.