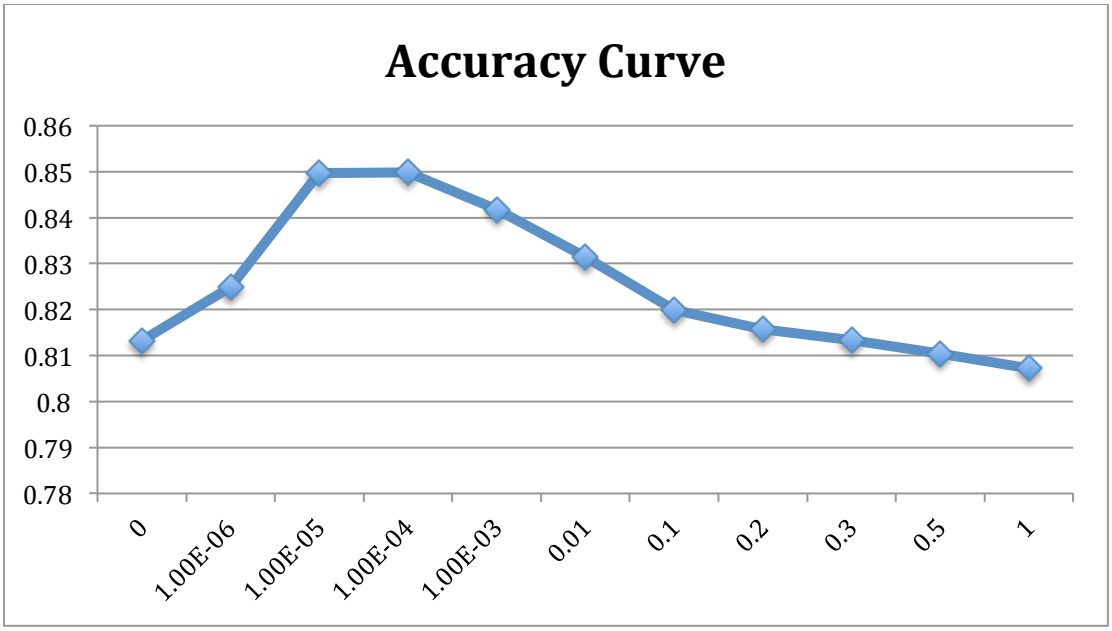
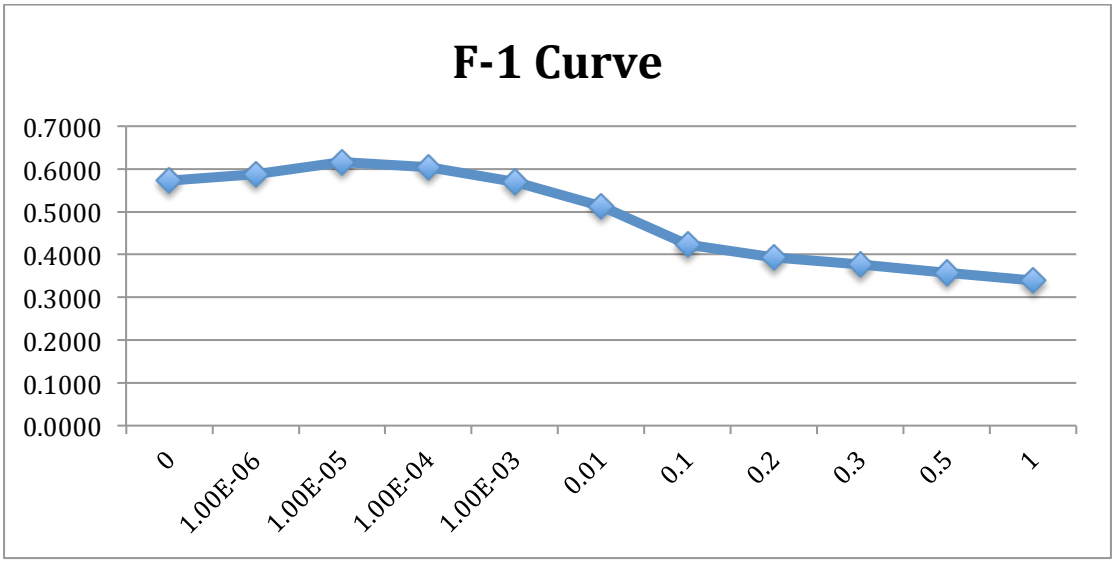
Homework 5
Daniel Ribeiro Silva
drsilva

1)
LCL at iteration 1 -> -228747.38303294237
LCL at iteration 2 -> -66168.5186525018
LCL at iteration 3 -> -30577.436480590834
LCL at iteration 4 -> -18333.595022200563
LCL at iteration 5 -> -12116.668850751586
LCL at iteration 6 -> -8550.177398312664
LCL at iteration 7 -> -6298.9634142478735
LCL at iteration 8 -> -4829.23224119037
LCL at iteration 9 -> -3855.342460812859
LCL at iteration 10 -> -3137.5657603932086
LCL at iteration 11 -> -2606.682976482325
LCL at iteration 12 -> -2183.983213182021
LCL at iteration 13 -> -1868.1330868686816
LCL at iteration 14 -> -1711.246761356121
LCL at iteration 15 -> -1324.0094623042794
LCL at iteration 16 -> -1211.5402310758723
LCL at iteration 17 -> -1105.8758062243567
LCL at iteration 18 -> -1077.845667801369
LCL at iteration 19 -> -830.7465240784657
LCL at iteration 20 -> -800.1341371748347

2) Vocabulary size: 100,000

μ	0	1e-6	1e-5	1e-4	1e-3	0.01	0.1	0.2	0.3	0.5	1
TP	493453	490046	475397	451797	411982	349659	259721	234819	222035	206851	194600
TN	2705567	2755202	2867480	2891718	2899868	2921667	2965641	2973982	2978057	2981436	2981019
FP	405110	355475	243197	218959	210809	189010	145036	136695	132620	129241	129658
FN	329660	333067	347716	371316	411131	473454	563392	588294	601078	616262	628513
Accu-racy	**0.8132**	**0.8249**	**0.8497**	**0.8499**	**0.8418**	**0.8315**	**0.8199**	**0.8157**	**0.8134**	**0.8104**	**0.8072**
Preci-sion	**0.5492**	**0.5796**	**0.6616**	**0.6736**	**0.6615**	**0.6491**	**0.6417**	**0.6321**	**0.6261**	**0.6155**	**0.6001**
Re-call	**0.5995**	**0.5954**	**0.5776**	**0.5489**	**0.5005**	**0.4248**	**0.3155**	**0.2853**	**0.2698**	**0.2513**	**0.2364**
F1	**0.5732**	**0.5874**	**0.6167**	**0.6049**	**0.5699**	**0.5135**	**0.4230**	**0.3931**	**0.3770**	**0.3569**	**0.3392**

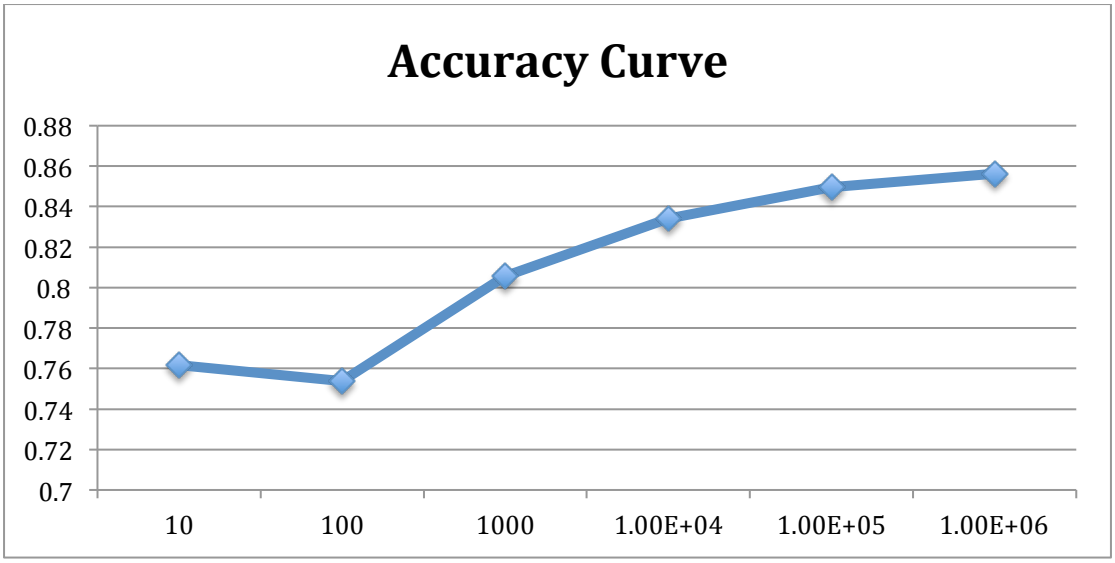


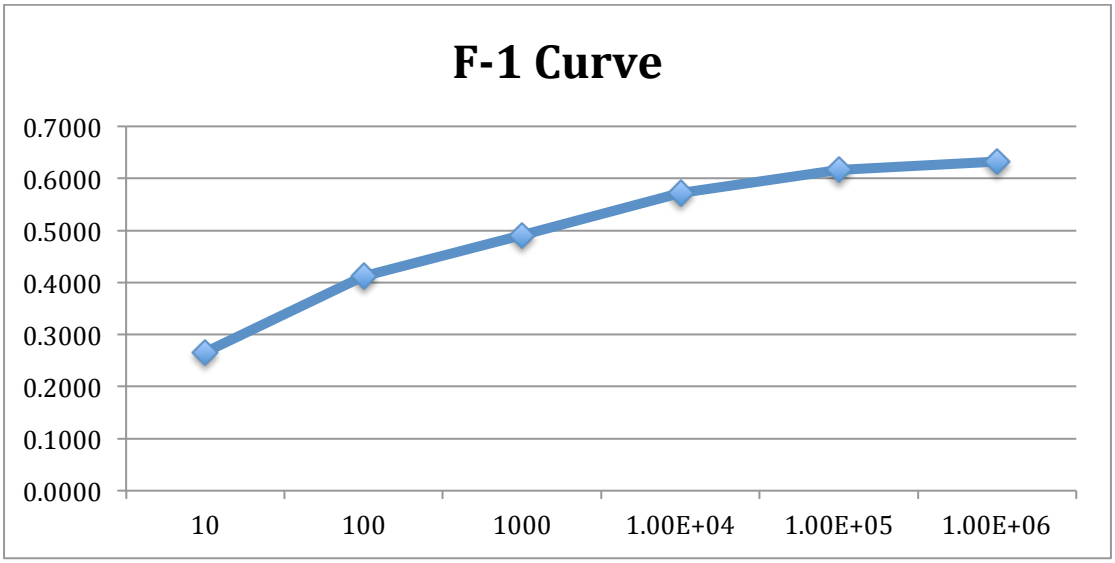


Considering both F1 score and Accuracy, I decided to choose $\mu = 1e - 5$.

3) $\mu = 1e - 5$

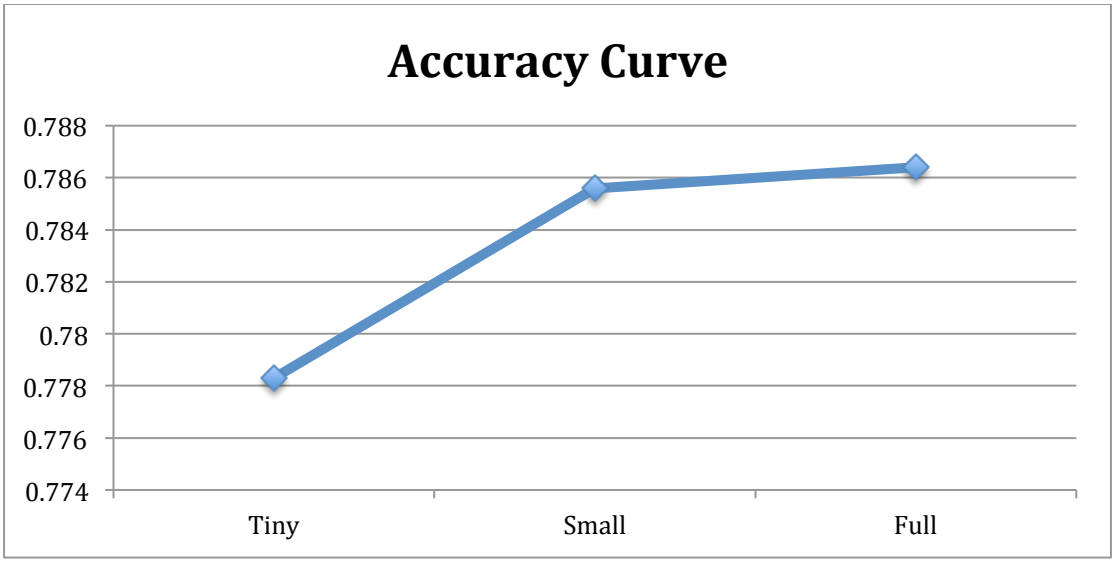
Vocab	10	100	1000	1e4	1e5	1e6
TP	169760	339402	368956	436815	475397	486152
TN	2826556	2625894	2800173	2845078	2867480	2882676
FP	284121	484783	310504	265599	243197	228001
FN	653353	483711	454157	386298	347716	336961
Accuracy	**0.7616**	**0.7538**	**0.8056**	**0.8342**	**0.8497**	**0.8563**
Precision	**0.3740**	**0.4118**	**0.5430**	**0.6219**	**0.6616**	**0.6807**
Recall	**0.2062**	**0.4123**	**0.4482**	**0.5307**	**0.5776**	**0.5906**
F1	**0.2659**	**0.4121**	**0.4911**	**0.5727**	**0.6167**	**0.6325**

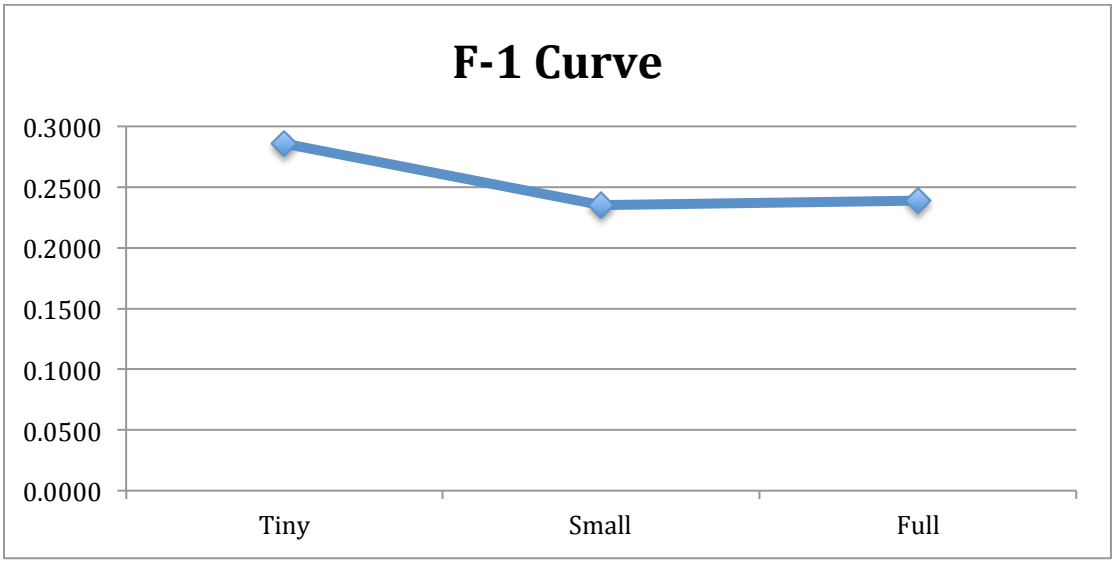




4) For this question, I used a vocabulary size of 10,000 for the tiny and small datasets and 100,000 for the full dataset.

μ	Tiny	Small	Full
TP	18	1278	131982
TN	298	29188	2961674
FP	11	1492	149003
FN	79	6822	691131
Accuracy	**0.7783**	**0.7856**	**0.7864**
Precision	**0.6207**	**0.4614**	**0.4697**
Recall	**0.1856**	**0.1578**	**0.1603**
F1	**0.2857**	**0.2351**	**0.2391**





5) Had brief discussions with Jeff Gee about using F1 as metric. Saloni Potdar and Poorna helped shuffled the data for me on their Linux computer.