10605 BigML Assignment 5:

Hao Gao (haog)

Did you receive any help whatsoever from anyone in solving this assignment? No

Did you give any help whatsoever to anyone in solving this assignment? No

Q1

Log likelihood output for small with dictionary size 10000 and μ =0.1:

-235751.6884086138

-129668.6646817663

-82615.64882866481

-74232.34308401159

-79339.44209480262

-71771.53538579625

-70580.47799691425

-71707.37024523772

-72867.84522933248

-71257.94863523961

-69078.08076103363

-70697.39974563085

-68465.81647070107

-70064.93401673773

-69939.81033543663

-70439.52871302719

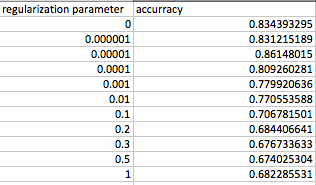
-69546.82161295884

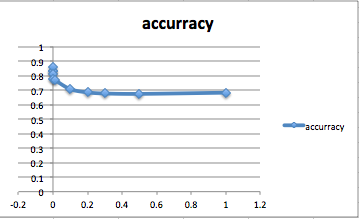
-69320.2698505788

-68435.2431052846

-68708.35421567242

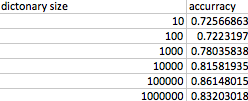
Q2

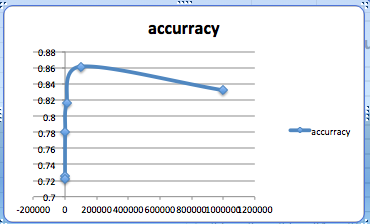




The best parameter is 0.00001.

Q3





The best dictionary size is 100000.

Q4

Treat the document as 14 documents for binary Naïve Bayes task for each of the 14 labels. Calculate the average accuracy.

|  |  |  |
| --- | --- | --- |
|  | NB | SGD |
| Tiny | 40.25 | 80.2 |
| Small | 80.25 | 80.65 |
| Full | 82.34 | 80.75 |