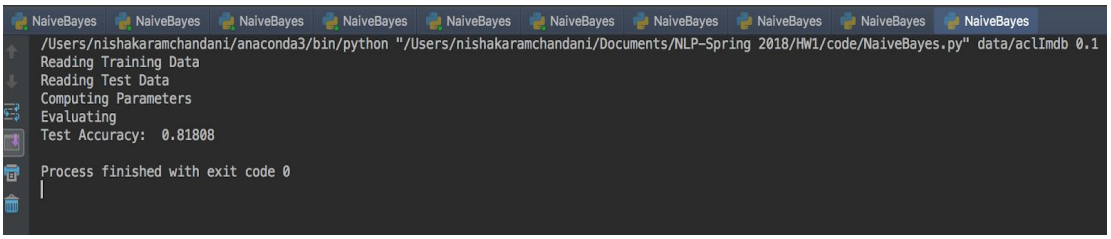
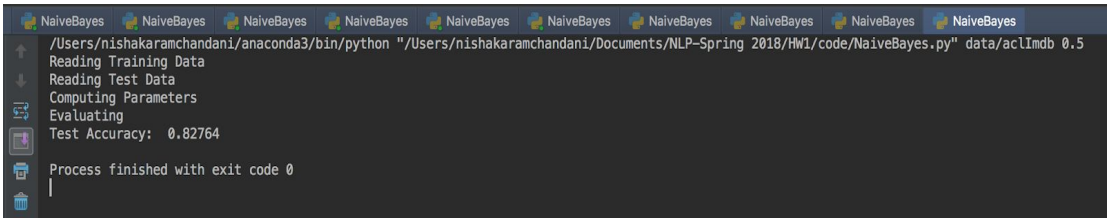
HW1 - Naive Bayes Text Classification
Nisha Karamchandani
A20414435

Classification and Evaluation:

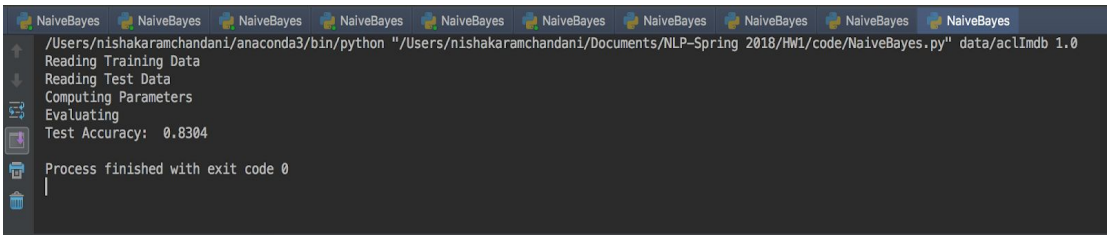
1. Alpha = 0.1. Test Accuracy: 0.81808



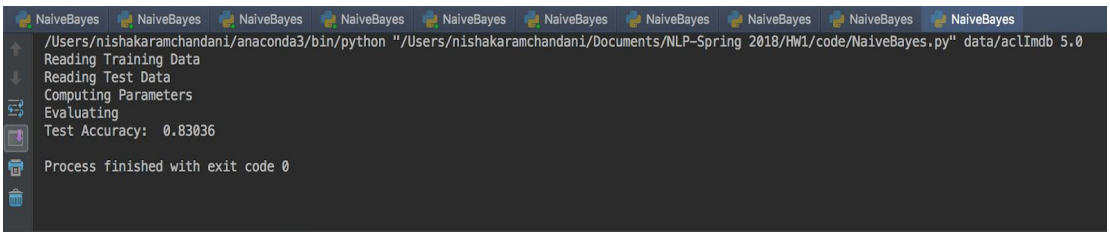
2. Alpha = 0.5. Test Accuracy: 0.82764



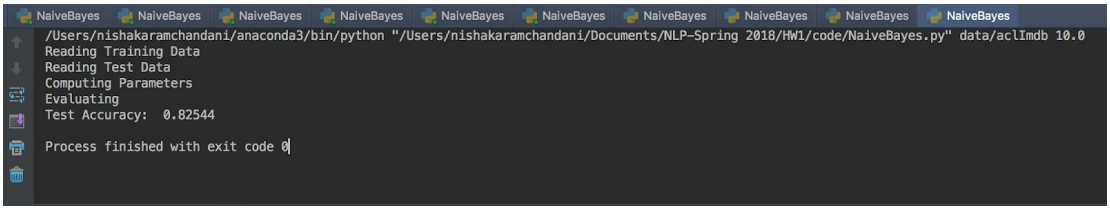
3. Alpha = 1.0. Test Accuracy: 0.8304



4. Alpha = 5.0. Test Accuracy: 0.83036



5. Alpha = 10.0. Test Accuracy: 0.82544

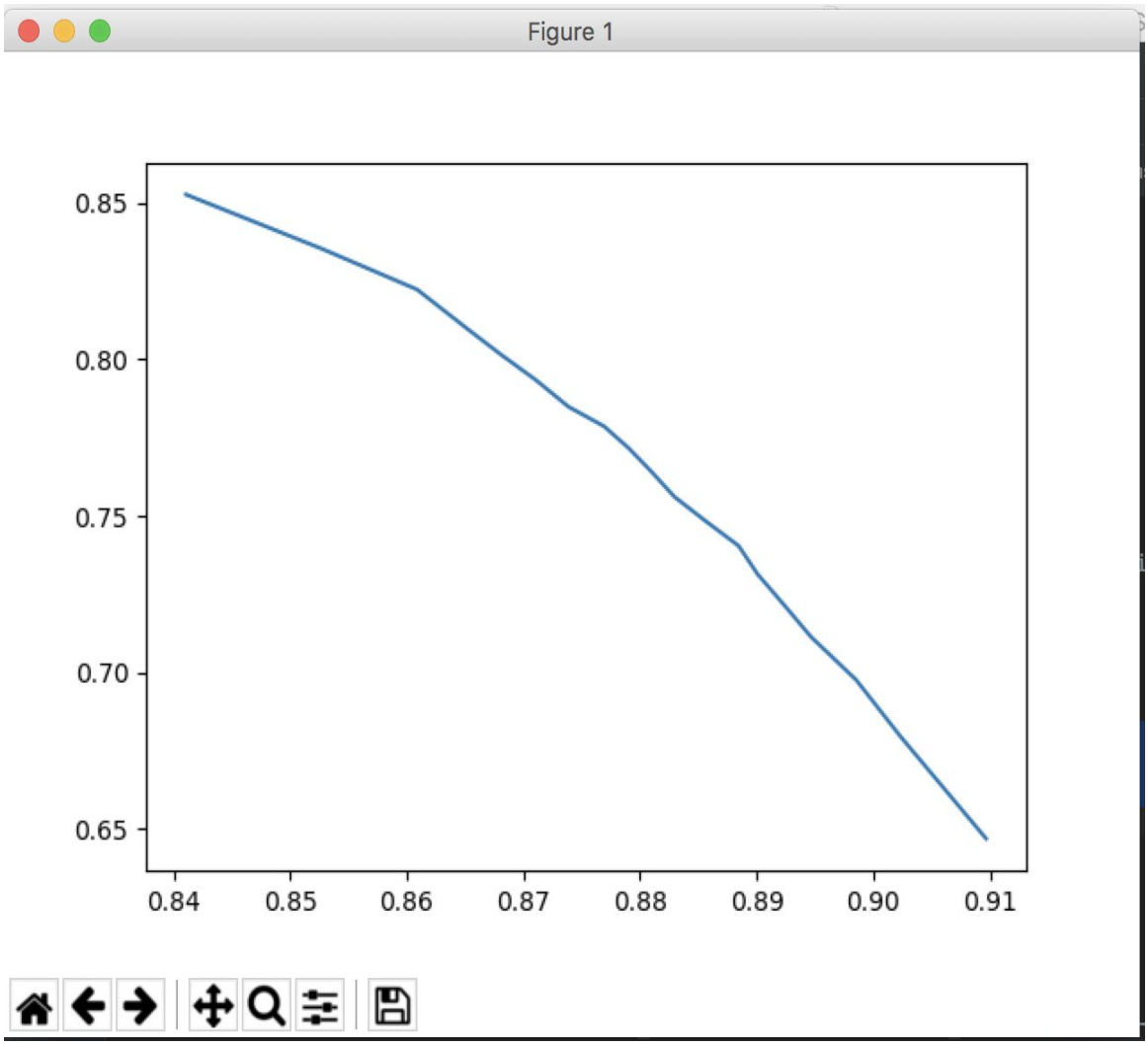


Probability Prediction:

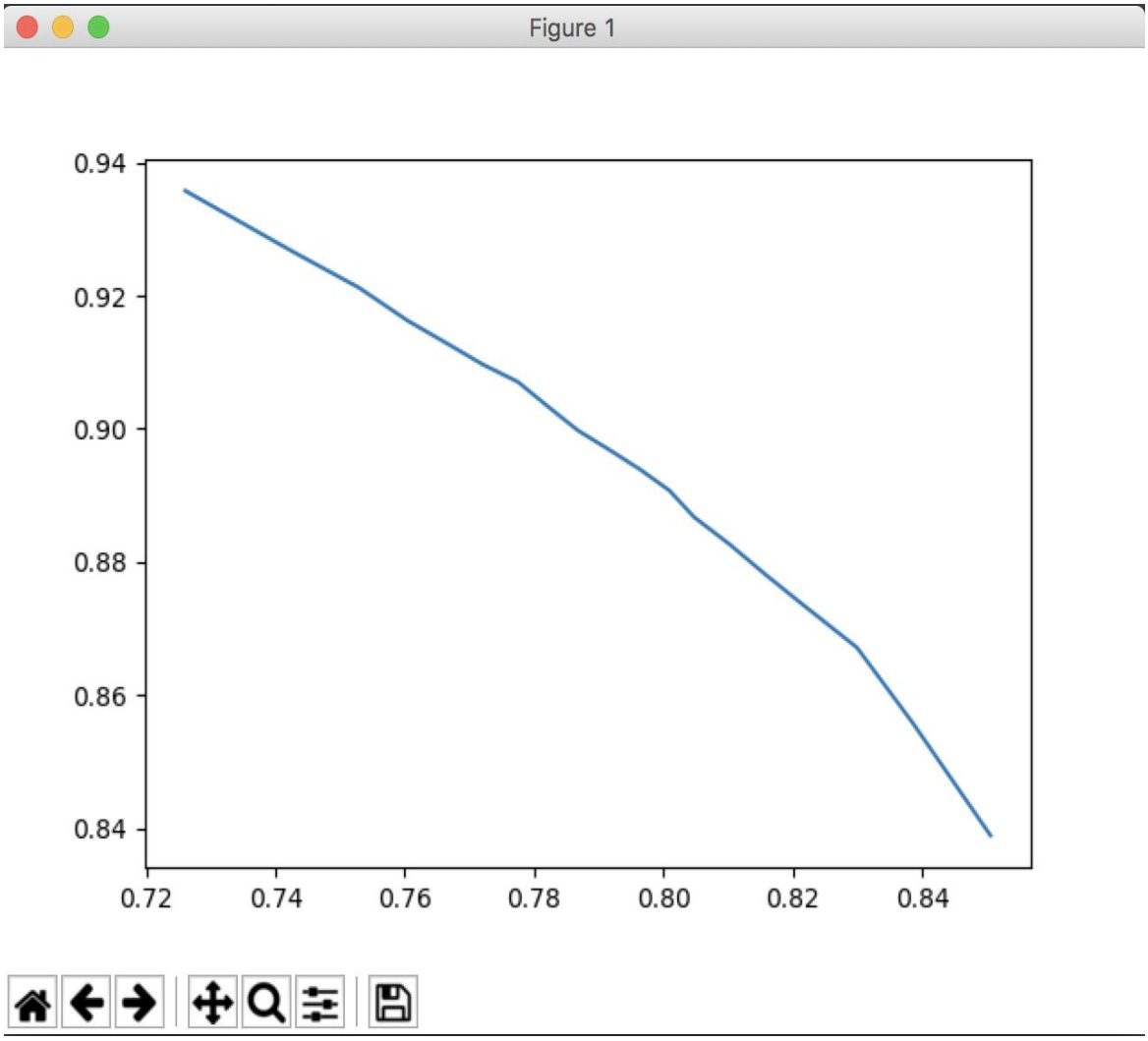
Calculated log probabilities for first 10 documents(Alpha = 1.0):
For Positive:
Calculated positive probability:
document_number:: 0 probability:: 1.5376009336806377e-11
document_number:: 1 probability:: 0.00030358367237154994
document_number:: 2 probability:: 0.029200350424162795
document_number:: 3 probability:: 1.5520759353285936e-09
document_number:: 4 probability:: 2.265542288881269e-06
document_number:: 5 probability:: 5.091173718956452e-06
document_number:: 6 probability:: 7.013386083803303e-17
document_number:: 7 probability:: 0.9978059708444659
document_number:: 8 probability:: 0.7241331060613385
document_number:: 9 probability:: 9.071686032112106e-11
For Negative:
Calculated negative probability:
document_number:: 0 probability:: 0.9999999999845386
document_number:: 1 probability:: 0.9996964163275854
document_number:: 2 probability:: 0.9707996495758465
document_number:: 3 probability:: 0.9999999984479473
document_number:: 4 probability:: 0.9999977344576307
document_number:: 5 probability:: 0.9999949088262381
document_number:: 6 probability:: 1.0
document_number:: 7 probability:: 0.0021940291554794268
document_number:: 8 probability:: 0.2758668939386897
document_number:: 9 probability:: 0.9999999999092779

Precision and Recall

1. Precision vs Recall for positive category:



2. Precision vs Recall for negative category:



The graph shows that there is always a trade off between the values for precision and recall.

Features:

Top 20 positive words and weights::

Word:: the weight:: 0.49984919075642165
Word:: and weight:: 0.4980288980364577
Word:: a weight:: 0.5128463011784161
Word:: of weight:: 0.5027198358386868
Word:: to weight:: 0.5051150339847803
Word:: is weight:: 0.5232330675521814
Word:: in weight:: 0.5017243316368926
Word:: i weight:: 0.5117516819184107
Word:: it weight:: 0.5059258556351555
Word:: that weight:: 0.5002913242193913
Word:: this weight:: 0.4971172649850061
Word:: as weight:: 0.4943312645284093
Word:: /><br weight:: 0.4763780681433915
Word:: with weight:: 0.4977801388273508
Word:: for weight:: 0.5104457852839301
Word:: was weight:: 0.5090572144436254
Word:: but weight:: 0.4627735272153555
Word:: his weight:: 0.5076414381856938
Word:: on weight:: 0.5006365231624912
Word:: film weight:: 0.49496415632561275

Top 20 negative words and weights::

Word:: the weight:: 0.5001508092435784
Word:: a weight:: 0.5019711019635423
Word:: and weight:: 0.4871536988215839
Word:: of weight:: 0.4972801641613132
Word:: to weight:: 0.49488496601521964
Word:: is weight:: 0.47676693244781865
Word:: in weight:: 0.4982756683631073
Word:: i weight:: 0.4882483180815893
Word:: this weight:: 0.4940741443648444
Word:: that weight:: 0.49970867578060874
Word:: it weight:: 0.5028827350149938
Word:: /><br weight:: 0.5056687354715907
Word:: was weight:: 0.5236219318566084
Word:: for weight:: 0.5022198611726493
Word:: but weight:: 0.4895542147160699
Word:: with weight:: 0.4909427855563746
Word:: as weight:: 0.5372264727846445
Word:: movie weight:: 0.4923585618143062
Word:: on weight:: 0.49936347683750876
Word:: not weight:: 0.5050358436743874