Homework 2

Author: Shashank Garg

Implement a majority class baseline

Data	Precision	Recall	F-score
Training	0.43275	1.0	0.604083057058105
Development	0.418	1.0	0.5895627644569816

Word Length Baseline

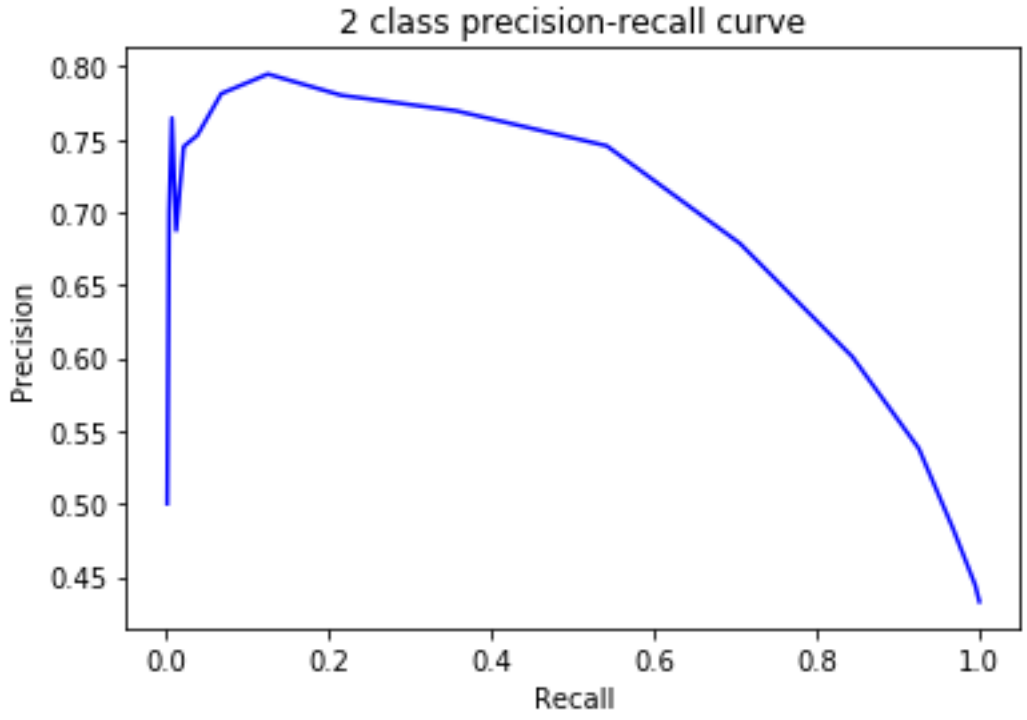
Data	Precision	Recall	F-score
Training	0.6007401315789473	0.8440207972270364	0.7018976699495555
Development	0.6053511705685619	0.8660287081339713	0.7125984251968505

Range of threshold tried

0 to max word length in the dataset provided

Best Threshold:

7



Word Frequency Baseline

Data	Precision	Recall	F-score
Training	0.5640718562874252	0.8162911611785095	0.6671388101983002

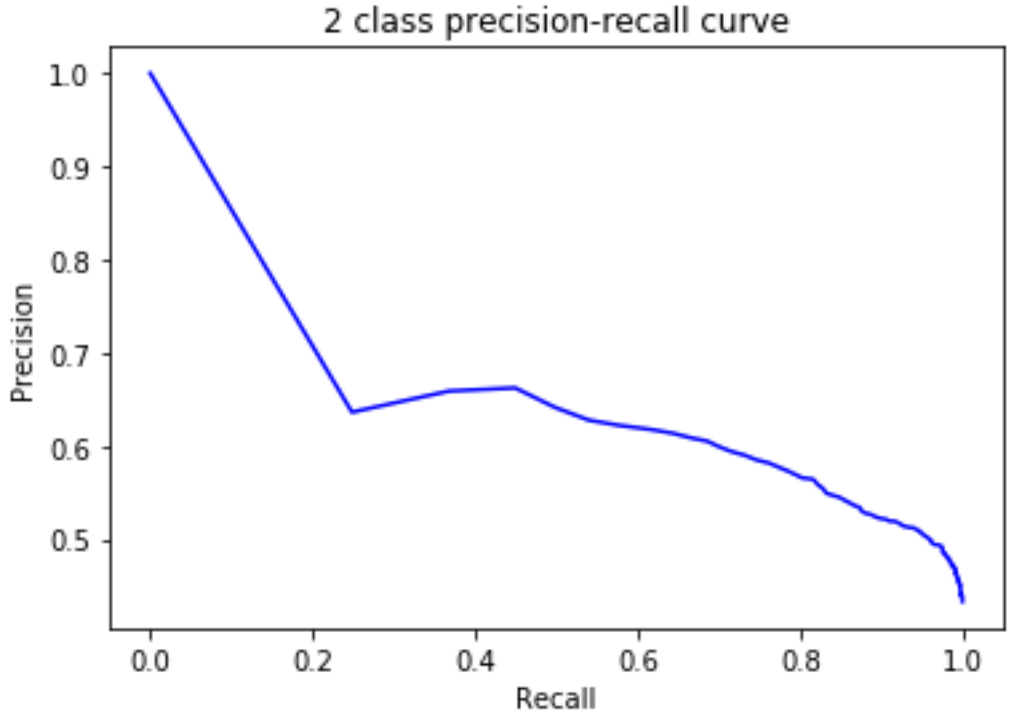
Development	0.5559055118110237	0.8444976076555024	0.6704653371320038

Range of threshold tried

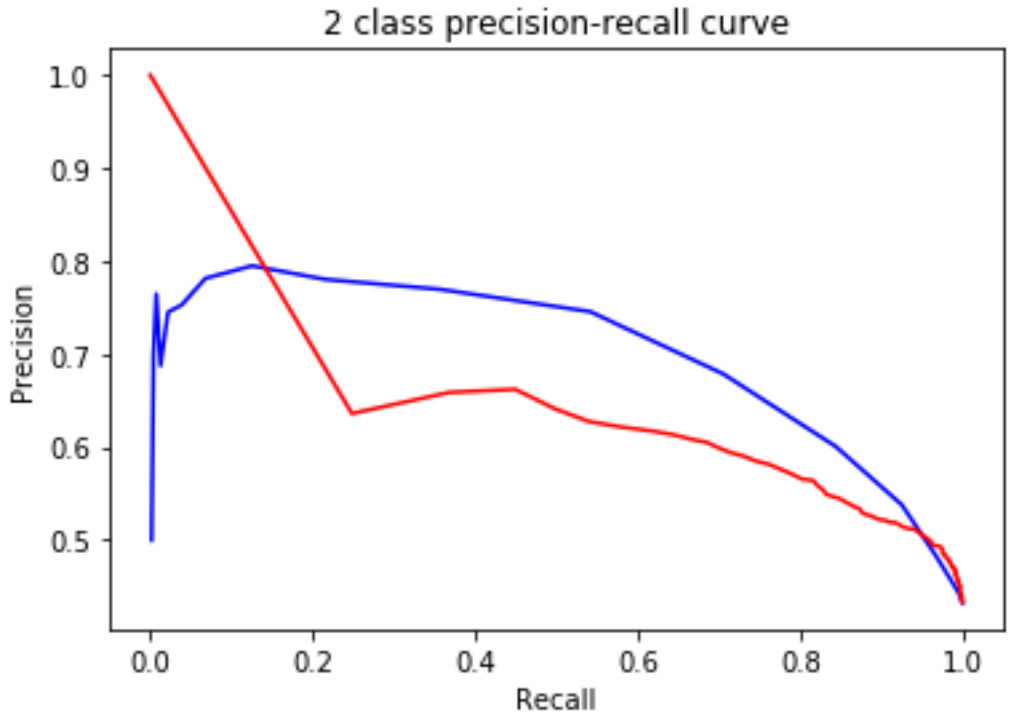
0 to 1e+10 with step size of 1e+7

Best Threshold

20000000



Comparing Word Length Baseline with Word Frequency Baseline



Red line indicates precision-recall curve for word frequency baseline model and blue line indicates precision-recall curve for word length baseline model.

Since area of the word length precision recall curve is greater than the area of word frequency precision recall curve, it can be suggested that the word length model is better than the word frequency model.

Naïve Bayes

Data	Precision	Recall	F-score
Training Data	0.4950379451255108	0.9797804737146159	0.6577467519875897
Development Data	0.46929316338354576	0.9688995215311005	0.6323185011709602

Logistic Regression

Data	Precision	Recall	F-score
Training Data	0.7250159134309357	0.6580011554015021	0.689884918231375
Development Data	0.7268170426065163	0.69377990430622	0.7099143206854344

The F-score of logistic regression model is higher than that of naïve bayes model for both the training data as well as development data. This indicates that the logistic regression model is expected to perform better than naïve bayes model on unseen data. One more observation to note from the results is that the recall of the naïve bayes model on both the training data as

well as the development data is close to 1. This indicates that the naïve bayes model predicts majority of the word observations as complex words which is thus taking a toll on its precision.

Build Your Own Model

Features Used

1. Word Length
2. Word Frequency as given in Google n-gram frequency count
3. Syllables
4. WordNet synonyms

Models Tried

- Random Forest
  I tried to tune the following hyperparameters by using grid search:
  - Max_depth
  - N_estimators

  I searched over the space of the depth of upto 12 and maximum esitmators of upto 1000. The best value was achieved at (max_depth, n_estimators)=(7,1000) and the following performance measure is reported:

Data	Precision	Recall	F-score
Training Data	0.8576086956521739	0.9116117850953206	0.8837860543265191
Development Data	0.7152173913043478	0.7870813397129187	0.7494305239179955

Examples Correctly Classified

- Unit
- Chairwoman
- Crippling
- speed
- color
- syndrome
- misbehavior
- form
- beige
- crowded
- proponents
- unraveled
- nutrition
- pull
- tanks
- troubles
- show
- objected
- desk
- government
- protein

- embedded
- gazing
- whisked
- believe
- shuddering
- safest
- twirled
- performance
- stirred
- unprecedented
- gave
- close
- visionary
- class
- feeling
- cousin
- wrong
- encourage
- targeting
- cattle
- told
- bald
- going
- purchasers
- princesses
- bristled
- secret
- overturned
- minds
- biomedical
- coach
- victims
- muscle
- epidemiologist
- snapshot
- talks
- promoted
- clinic
- games
- left
- 4-May
- proposal
- worsened
- astronomer
- portable
- otherworldly
- commotion
- promoting

- narrating
- next
- incentives

Examples Incorrectly Classified

- practically
- perpetrated
- patterns
- legacy
- elevator
- daughters
- majority
- beaver
- embrace
- cleanup
- eerie
- helicopter
- cookie
- behaviors
- pups
- attorney
- politicians

Categories incorrectly classified

- Words whose length was in the range of 5-8 characters were misclassified more often
- Very long words such as self-imposed, elementary were also misclassified more.

Optional Data Used

SemEval 2016 Dataset: http://alt.qcri.org/semeval2016/task11/