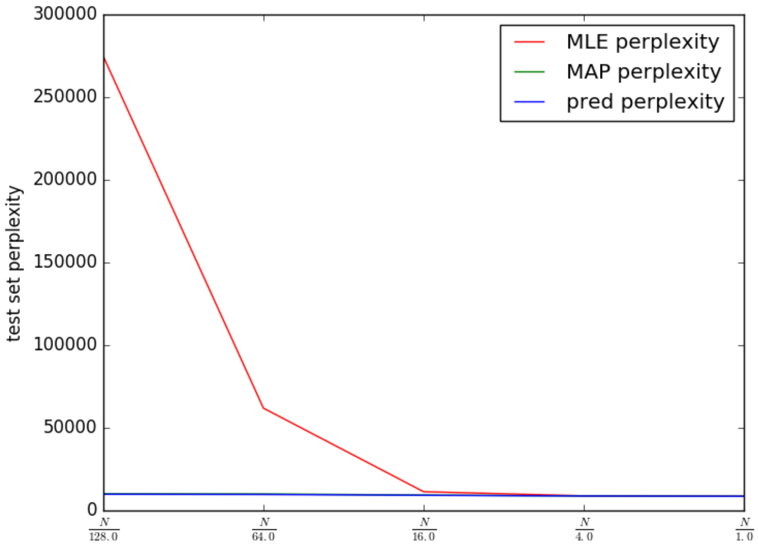
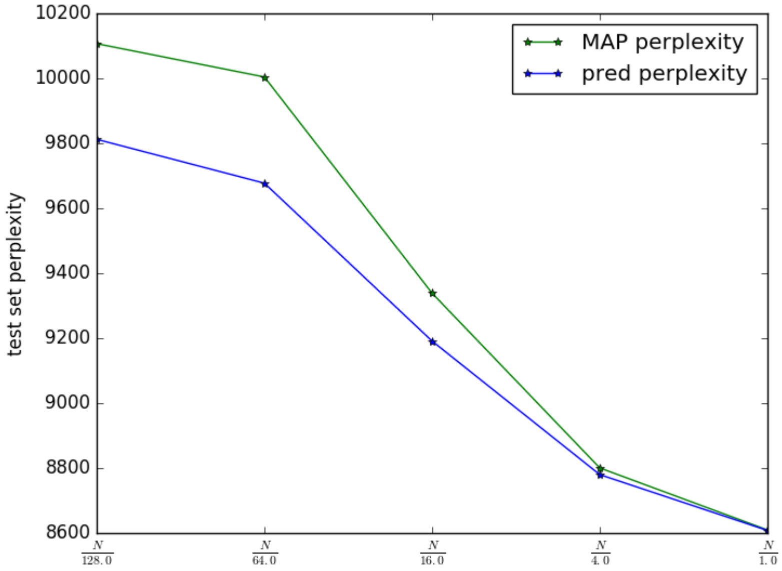
COMP136 Project_1 Test Report

By beibei du

Task 1: Model Training, Prediction & Evaluation

- What happens to the test set perplexities of the different methods with respect to each other as the training set size increases? Please explain why this occurs.





According to figure above, we can see that as the training set increases, all of three perplexities will decrease except for the maximum likelihood estimate.

For the maximum likelihood estimate, if we test a word that is not included in the train set or haven’t been included in the model, then we will get probability of this word is zero, then the perplexity will be infinity.

For the MAP and Predictive distribution model, we use the parameter alpha to work as the prior, which can give some value for the words in vocab, so when we test data, the probability won’t be zero. Since they both decrease as the training data size increases, this is maybe the alpha weighted less as more related words were involved in the training set . Also, we can find that MAP performance better than predictive distribution as the train data set is smaller. This is because it gives a weighted prediction for all models.

- What is the obvious shortcoming of the maximum likelihood estimate for a unigram model? How do the other two approaches address this issue?

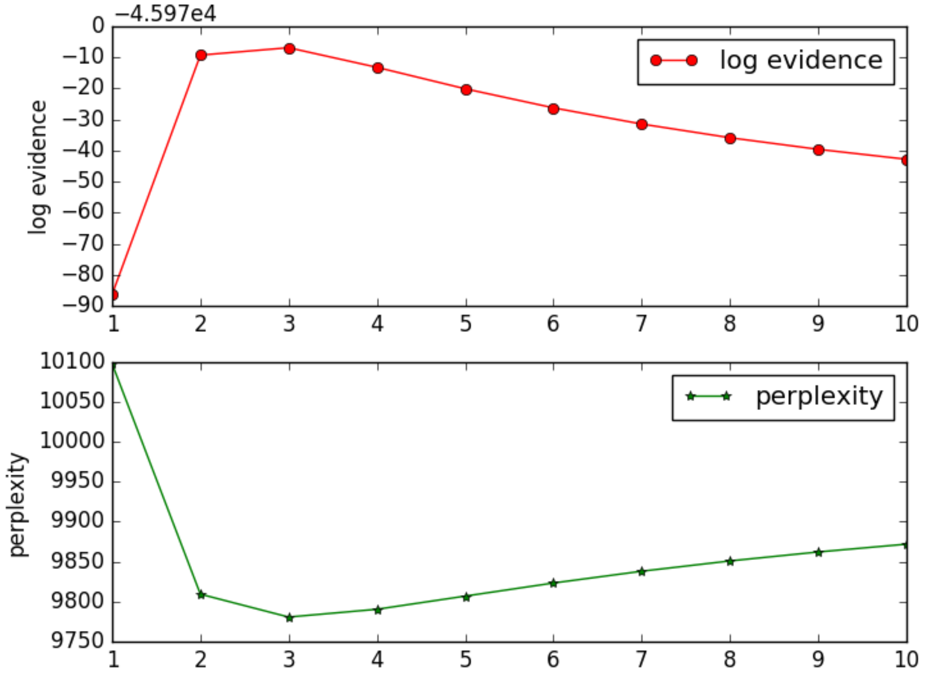
  For the maximum likelihood estimate, when the test data (e.g. w_i) is not included in the train set, then probability of w_i will be zero, so when we use perplexity formula to calculate the perplexity, the ln(p(w_i)) will be ln(0)->infinity.
  The other approaches using a parameter alpha to avoid this kind of situation. In the MAP estimate , mk + ak – 1, it at least is 1, so the probability won’t be zero. Same as predictive distribution, also use alpha to work as a prior to avoid this.

- For the full training set, how sensitive do you think the test set perplexity will be to small changes to a’?

  I think for the full training set, test set perplexity will be less sensitive the small changes to a’.

Task 2: Model Selection

- Is maximizing the evidence function a good method for model selection on this dataset?



According to the figure, we can find that when a’ = 3, the evidence will be in maximum and the perplexity will be in minimum. Which seems that maximize evidence can be a method for model selection in this dataset.

Task 3: Author Identification

- Was the model successful in this classification task?

According to the perplexity result.

We use pg84.txt as train data set, pg345.txt and pg118.txt as the test set, and get the below result:

```
    perplexity using pg84 txt file is 8270.715157
    perplexity using pg1188 txt file is 5864.369457
```

The lower perplexity with name ‘pg1188’ will have same author as pg84.So we can conclude that this model is successful in classification task.