# Assignment 2: Maximum Entropy Classification

Meg Kobashi

October 2017

## 1 Approach

The model was tested on the names corpus, to classify by gender, and the Yelp reviews corpus, to classify by "positive", "negative", "neutral." After adjusting the parameters for all of the mini-batches, we perform classification on the development instances, and record the accuracy and negative log-likelihood. The size of the development set is 1000 instances across all runs. If the accuracy does not improve for at least 4 iterations, we determine that it has converged. The highest accuracy on the test set was 76% for the names corpus and 69% for the reviews corpus.

## 2 Experiments

The charts are on the last page!

### 2.1 Experiment 1

The size of the mini-batch size {1, 10, 30, 50, 100, 1000} is plotted against their respective accuracy on the development set. The mini-batch size in the range 1-100 yielded the best results; there was not much of a different in accuracy among batch sizes in this range. However, a batch size of 1000 yielded an accuracy of 45% for the reviews corpus, which was significantly lower than the smaller batch sizes.

### 2.2 Experiment 2

The size of the training set {1000, 10000, 20000, 50000, 100000} is plotted against their respective accuracy on the development set. The accuracy increases as we increase the size of the training set; however, it drops once we reach a peak at 20000 instances.

# 3 Results

As a result of experimentation described in the previous sections, we found that the following experiment setting yielded the best results:

- Learning rate: 0.01
- Mini-batch size: 10, 30, or 50
- Training set size: 20,000
- Features: Most common 1000 words out of all words in the training set

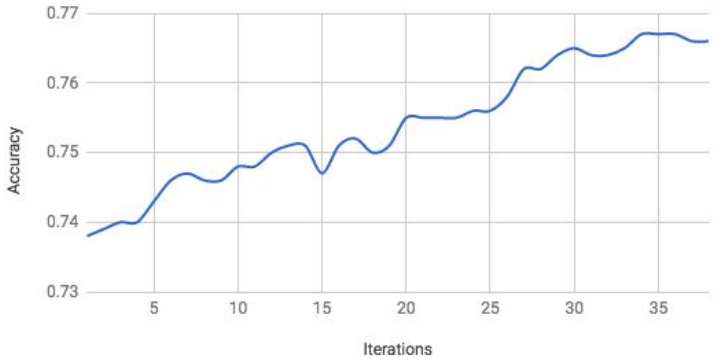
Increasing the number of features did not improve the accuracy by much, but only increased the runtime for training. Therefore, it was better to limit the feature set to 1000 words for the sake of performance.

# 4 Future Improvements

- Improve runtime by speeding up calculations using numpy matrices.
- Experiment with different conditions for convergence; for example, using negative log-likelihood instead of accuracy.

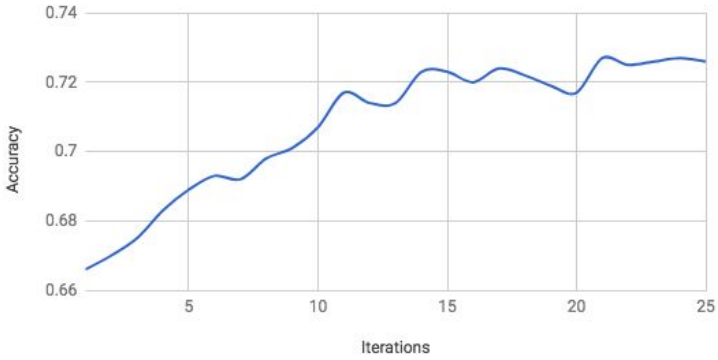
**Name classification**

**Batch size: 30**



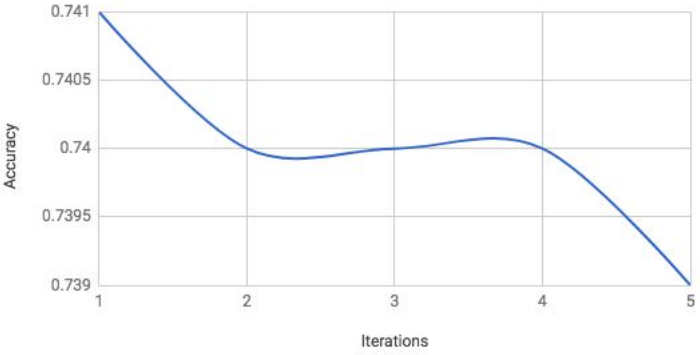

**Review classification using 1000 most common features**

**Batch size: 30**



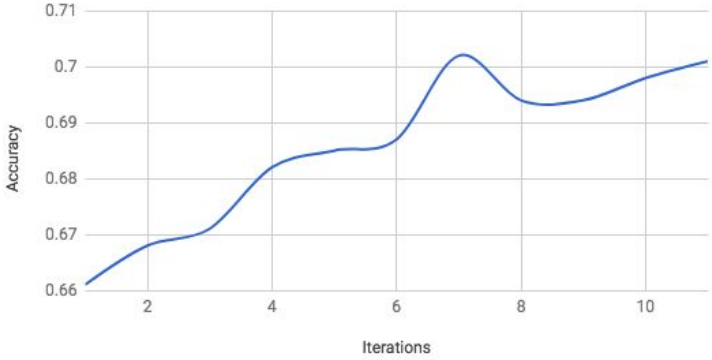

**Name classification**

**Batch Size = 100**



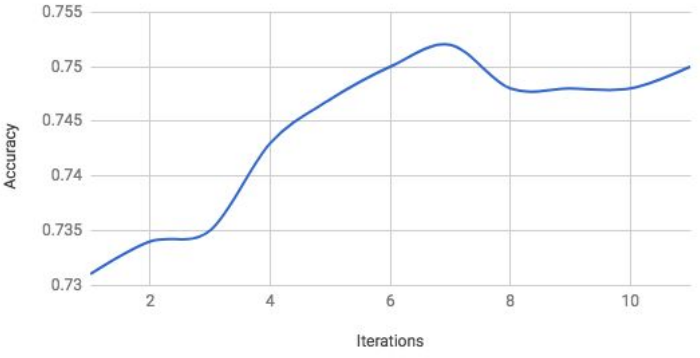

**Review classification using 1000 most common features**

**Batch size = 100**



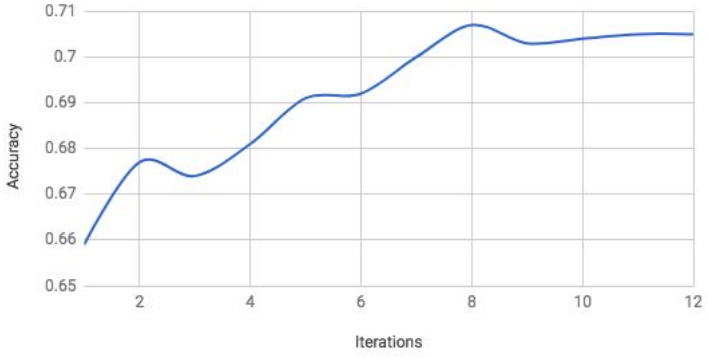

**Names Classification**

**Batch size = 1**



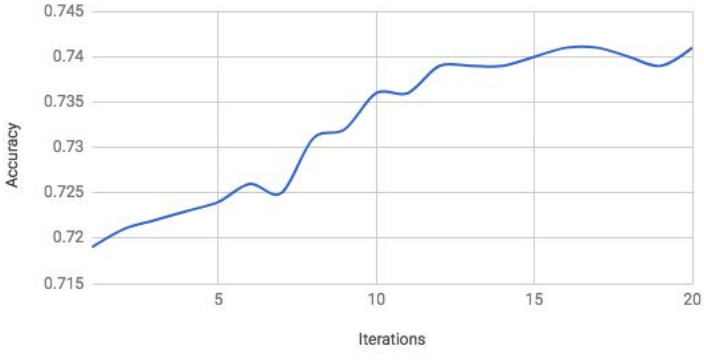

**Review Classification using 1000 most common features**

**Batch size = 1**



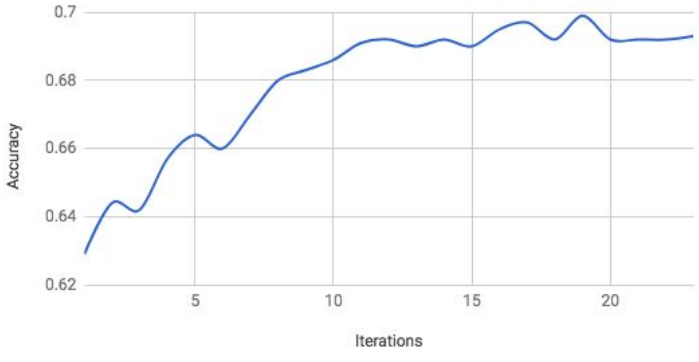

**Names Classification**

**Batch size = 10**



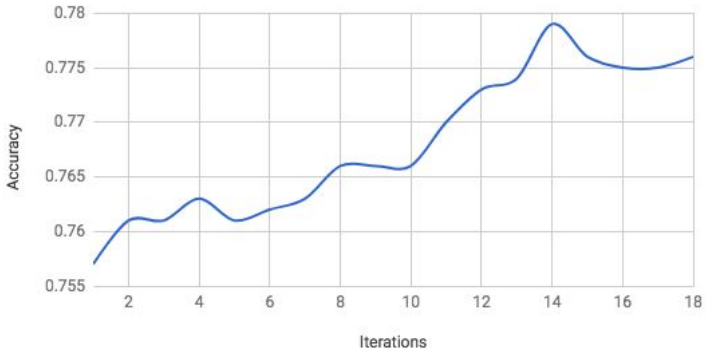

**Review Classification using 1000 most common features**

**Batch size = 10**



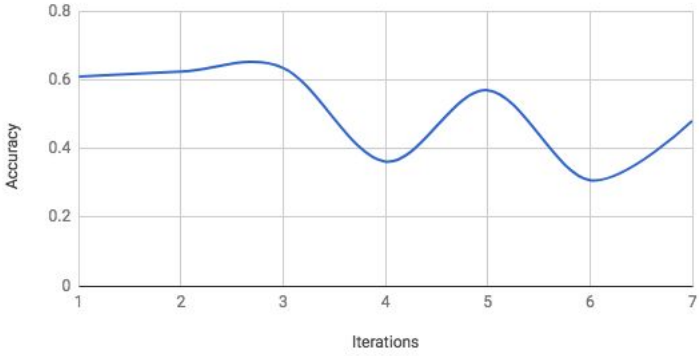
**Name classification**

**Batch size = 1000**




**Review classification using 1000 most common features**

**Batch size = 1000**




Effect of Batch Size on Maximum Accuracy