Machine learning for computational linguistics: Assignment 3

Fernando Ardente

August 1, 2016

1 logistic regression classifier

Acc: 0.86799999999999999

2 mlp

Acc (1000 hidden units): 0.62149999999999994

3 CNN

Acc: 0.4950

4

Somehow, the logistic regression has the highest accuracy, even though the CNN was expected to to at least similarly well. I guess, using the pretrained GLOVE Vectors might help, instead of training the vectors only on our small dataset. Also, playing around with the parameters was easier with the logistic regression classifier, especially since the CNN takes quiet some time to converge on my poor laptop. The multilayer perceptron did better than I would have expected,since there were a lot of features involved. It also finishes faster than the CNN, but depending on the hidden unit size and the features used, it can also get my machine to crash. For the logistic regression classifier, I found that Bigrams actually performed best,together with keeping words with frequencies between 2 and 200. Also keeping stopwords seemed to have a positive effect (although it is not really sure they were kept due to the limit of 200 occurrences).