

Assignment 3

Arie Cattan 336319314
Moche Uzan F957022

1. Can the two languages be distinguished using a bag-of-words approach? Explain why.

No, they can't because the bag-of-words approach does not keep the order, and in our case, positive and negative classes have the same bag of word.

2. Can the two languages be distinguished using a bigram or trigram based approach? Explain why.

Apparently no, because even though these models consider the order unlike the CBOW approach, we do not care about exact position in the sequence, and in our case, the difference between positive and negative samples is exactly the position.

But, with a simple preprocessing on the data, which consists to drop the digits characters and to represent a letter sequence by only one letter, bigram and trigram models should work.

3. Can the two languages be distinguished using a convolutional neural net- work? Explain why.

Yes, because the conv layer will detect which n-grams are predictive.