בס"ד

STEVE GUTFREUND – 342873791
YOEL JASNER - 204380992

Can the two languages be distinguished using a bag-of-words approach?
No. A bag-of-words model is disregarding the word order. The language defined in our assignment is mainly based on order (‘b’ **before** ‘c’ and there must be digits **before and after** each letter.
An advantage of ‘bow’ is that it remembers the quantity of each word, but in our language it’s useless, we just care about order and not quantity (well we do need to know there’s at least one digit, one ‘a’ etc.)
A simple example why BOW won’t work is: 6a6b6c6d6 VS 6a6c6b6d6, a BOW based model would approach them as equal.

Can the two languages be distinguished using a bigram or trigram based approach?
No. The use of bigrams, trigrams or ngrams in general is good when in the specific language those parts appear in some distribution. In our case, for example, a digits sequence [1-9]+ can totally be random. If the duo ‘34’ appears 20 times or one time, doesn’t tell us anything about the validity.

Can the two language be distinguished using a convolutional neural network?
No. The main property of a CNN model are the convolutional layer(s), where there is ‘window’ sliding over the picture and processes every part. It does not per se remember the parts of the input. Our language in mainly based on the order of the inputs.