

Deep Learning - 89687

Ass3 – Part 1 - Challenge

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

No, since the bag-of-words method does not maintain the order. The window does not carry information about the order of the characters in the strings but the syntax form of the language requires order.

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

It's true that these two based approaches consider the order but they don't consider the exact position in the sequence. In our exercise, we can only differentiate the two types of examples through the position.

For example, 1a2c33b44d5. Bigram will take into consideration '2c' and 'b4', while trigram will see 'c33' and 'b44'. The order of 'c' and 'b' is missing.

So, our response is No.

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

Yes. A CNN can classify the two languages as it learns features that are related to the order of the language (with the help of the filters).