Natural Language Programming

Quiz 2

Thursday, 22 Sep 2022
Time: 10 mins
Marks: 10

Note: Please write in blue or black ink. Use of pencils is not allowed. No articles (calculators etc.) can be exchanged.

1. Write the Bayes' theorem in terms of classification.

   Solution: $p(class/document) = \frac{p(document/class) * p(class)}{p(document)}$

2. Write the probability of the sentence "Intelligence skips a generation" by taking the simplifying assumption that any word is at best dependent on the previous word.

   Solution: $p(Intelligence) * p(skips/Intelligence) * p(a/skips) * p(generation/a)$

3. Consider the corpus given below:
   <s> Classification is the task of categorizing instances based on their features. </s>
   <s> Classification can be tricky when features are not well defined. </s>
   <s> Currently both deep and hand crafted features are used to identify one class from the other. </s>

   Find the following probabilities:

$p(classification)$		$p(classification/< s >)$		$p(classification, < s >)$	

   Note: Show your working to get credit and state assumptions if any.

   Assumptions: 1) The period '.' is considered to be a separate token.
   2) <s> and </s> are also considered as separate tokens.

$$p(classification) = \frac{number\ of\ times\ 'classification' occurs\ in\ the\ corpus}{size\ of\ the\ corpus} = \frac{2}{46}$$
$$= 0.04347826$$

$$p(classification/< s >) = ?$$

Since it assumes that the previous token is the start of sentence marker, we only need to count the number of sentences in the corpus which is 3. Next we need to see how many

times a start of sentence marker is followed by ‘classification’, we can see that among the 3 sentences 2 of them start with ‘classification, and hence the required probability is $\frac{2}{3}$.

$$p(classification, < s >) = ?$$

To calculate the joint probability, we need to use the expression for joint probability in terms of conditional probability: 1) $p(classification, < s >) = p(classification/< s >) * p(< s >)$; 2) $p(classification, < s >) = p(< s >/classification) * p(classification)$. But since we have already figured out $p(classification/< s >)$, we shall use the 1st expression. Hence, the required probability is: $\frac{2}{3} * \frac{3}{46} = \frac{2}{46} = 0.04347826$.