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. What is the likelihood in the Bayes' theorem? Give an example of it from the real world.

   Solution: The term that estimates the probability distribution of the data (document in NLP) given the class is the likelihood term i.e. $p(d/c)$. It estimates the probability of observing a particular event given that a particular other event has been observed to be true. The examples discussed in the class pertain to: the probability of observing high humidity when its raining; the probability of observing the phrase "interest rate" in a document belonging to class 'sports'; the probability of observing MRI being positive of a patient who has cancer; or the probability of observing the finger prints of a person who has carried out the theft.

2. Write the probability of the sentence "I prefer mathematics over coding" without taking the Markov assumption.

   Solution: $p(I) * p(prefer/I) * p(mathematics/I\ prefer) * p(over/I\ prefer\ mathematics) * p(coding/I\ prefer\ mathematics\ over)$

3. Consider the corpus given below:
   <s> The movement of the tectonic plates give rise to earthquakes. </s>
   <s> A few of these earthquakes can wreak havoc on life and property. </s>
   <s> In the past a few of these earthquakes measuring almost 10 on the Richter scale have caused the downfall of civilizations. </s>

   Find the following probabilities:

$p(earthquakes)$		$p(scale/Richter)$		$p(Richter, scale)$	

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

   Assuming that we are only considering lemmas after lemmatization, so 'earthquakes' would become 'earthquake'. The period '.' is counted as a separate token. <s> and </s> would be counted as separate tokens.

$$p(earthquakes) = \frac{number\ of\ times\ lemma\ of\ earthquakes\ occur\ in\ the\ corpus}{number\ of\ words\ in\ the\ corpus}$$
$$= \frac{3}{52} = 0.05769$$

$$p(scale/Richter) = ?$$

To figure out we have to see how many times the word 'Richter' occurs in the corpus and how many of those times it is followed by 'scale'. We can see that the word 'Richter' only occurs once and it is followed by 'scale' thus given the data it can be assumed that the same pattern would be observed in all the future occurrences of 'Richter' and hence its probability is 1.

$$p(Richter, scale) = ?$$

Applying the chain rule the joint probability can be written as a product of conditional probability and the marginal probability, hence, the above expression would become $p(Richter, scale) = p(scale/Richter) * p(Richter) = 1 * \frac{1}{52} = 0.01923$. The joint probability given above could also be rearranged as $p(Richter, scale) = p(Richter/scale) * p(scale)$, but since we have already calculated $p(scale/Richter)$, it would be easier to use the first expression.