LANGUAGE MODELLING (2 points / 20 minutes)

1. What is a language model, that is, what is it's purpose in terms of probabilities ? (0.5 points)

> A language model is an algorithm / method / formula to assign a probability to a sentence or to a sequence of words, in a given natural language. For instance, $P (\text{~~Avui fa bon dia~~}) = 0.0123$, and also, the conditional probability of a certain word given one or more previous words $P (\text{temps | Avui fa bon}) = 0.3$

2. If $w_{k-1} w_k$ is a bigram, which is the meaning of this equation ?

$$P(w_k | w_{k-1}) = \displaystyle\frac{C(w_{k-1} w_k)}{\sum_{w \in V} C(w_{k-1} w)} = \displaystyle\frac{C(w_{k-1} w_k)}{C(w_{k-1})}$$

- Explain what are V, w_k, w_{k-1}, how to compute $C(w_{k-1}), C(w_{k-1} w_k)$, and the meaning of probability being computed. (0.5 points)

> V is the vocabulari, set of all possible words. w_k, w_{k-1} are two consecutive words. C means count n-grams = number of occurrences of a certain sequence of one or more words. It's a conditional probability : prob. that next word is w_k given that previous word in a sentence is w_k.

3. One of the applications of a language model is spell checking. We have seen there are two types of spell checking, named *real word* and *non-word*. What is the difference, which one is harder and why ? (0.5 points)

> non-word spelling means to spot the typed words that do not belong to the vocabulary = permissible words, and offer the user a sorted list of possible right words in place of it. They are at most at edit distance 1 or 2 from the typed word. And sorted by probability according to the language model.
>
> Instead, in real-word spelling we have to detect mispelled words that belong to the vocabulary but are wrong because of its context: `She hat lived happily`. It is more difficult because each typed word is a potential mistake.

4. How can a language model help in real word spelling correction ? In order to answer it, first explain the meaning of the problem formulation

$$\argmax_{W \in C(X)}{P(X | W) P(W)}$$

- In particular, what is $C(X)$ and how is it computed ? Explain which one of the two probabilities is computed by a language model. Hint: X, W are sentences and X is the typed one, $P(W), P(X|W)$ are the prior and likelihood, respectively. (0.5 points)

> $C(X)$ is the set of candidate sentences for which to evaluate the probability. It results from assuming X has at most 1 mispelled word at edit distance at most 1 from the type word, for each word in X. W is some sentence in $C(X)$. The language model computes the prior probability for each W.

Validation questions for the language model exercise

1. In the first part of the exercise you were asked to try three variants of n-gram language models. Which one was not one of them ? a) MLE b) EveryGram c) Laplace d) StupidBackoff

> EveryGram

2. Second part of the exercise was about training a small feedforward neural network to learn a language model. Once we had trained the network we could generate sentences word by word, by feeding the network with the last few generated words. The network output was then, for each word in the vocabulary, the probability of being the next word, and we had to choose one. Two strategies were possible, to be implemented with numpy's argmax and torch.multinomial. Brielfy explain the idea of these two strategies, in one or two lines each.

> argmax : select the word with the highest probability. torch.multinomial : consider the probababilities for each word in the vocabulary as a distribution probability and draw a sample from it, therefore not necessarily the one with highest probability.

3. Last part of the exercise was about real word spell correction with a language model. Was this an n-gram model like in part 1 or a model learned by the neural network of part 2 ?

> It was a n-gram language model.