

**Natural Language Processing  
Final Exam**

September 23rd, 2022

1. **[2 points]** State the so-called distributional hypothesis, and explain when and how it is used in natural language processing.
2. **[5 points]** With reference to 2-gram models, answer the following questions.
  - (a) Introduce the basic idea underlying Laplace smoothing, and provide its mathematical definition.
  - (b) Reformulate Laplace smoothing using the notion of adjusted count  $C^*(w_t | w_{t-1})$ , and define the notion of relative discount.
3. **[6 points]** Consider the following term-context matrix, providing co-occurrences for target words car, book, airport and library against contexts words school, travel, sport and movies.

	school	travel	sport	movies
car	15	105	27	8
book	107	22	13	12
airport	5	87	0	3
library	103	0	0	2

For all entries in the main diagonal, indicate how to compute the positive pointwise mutual information. Use fractions and logarithms in your answers **without** computing these operators. For the purpose of this exercise, assume that no other word/context pairs matter, and assume that each context word appears in the context of only one occurrence of a target word, and the other way around.

4. **[2 points]** Describe the notions of extrinsic and intrinsic evaluations. Discuss the disadvantages of extrinsic evaluation.

(see next page)

5. **[5 points]** With reference to the contextualized language models, answer the following questions.
- Introduce the basic architecture of BERT.
  - Define and motivate the two training methodologies of masked language modeling and next sentence prediction.
6. **[5 points]** In the context of transition-based dependency parsing, consider the French sentence ‘Une lettre avait été envoyée la semaine dernière aux salariés’ along with the projective dependency tree consisting of the following unlabeled dependency relations.

<b>head</b>	lettre	envoyée	envoyée	envoyée	⟨ROOT⟩	semaine	envoyée	semaine	envoyée	aux
<b>dependent</b>	Une	lettre	avait	été	envoyée	la	semaine	dernière	aux	salariés

Answer the following questions.

- Draw a graphical representation of the dependency tree above, with arcs directed from the head to the dependent.
  - Apply to the above tree the oracle presented in class to construct a sequence of training instances for the arc-standard parser.
7. **[6 points]** Consider the end-to-end application of text-based question answering, and answer the following questions.
- Introduce the machine reading task and the notions of query, passage and span. Define the span probability  $P(p_i, \dots, p_j \mid q, p)$  and explain how this probability is approximated by means of start and end probabilities.
  - Introduce the Stanford attentive reader for machine reading, and discuss the main representation and equations used by this neural model.
8. **[2 points]** With reference to the evaluation of language models, define and discuss the perplexity measure.