IN4080, 2022, Exercise set 3: Sept. 27

In this course, you are tested in two different ways. The mandatory assignments test your skills, whether you are able to apply NLP. Much of the group/lab sessios also focus on this. At the exam, on the other hand, your theoretical understanding will be tested. To prepare you for the final exam, some of the (time of the) group sessions will be allocated to exam-like questions. The following questions are questions of the type you will meet at the exam.

Exercise 1

Assume we are classifying sentences into two classes, sentiment bearing and non-sentiment bearing and get the following confusion matrix.

		Gold	
		sentiment	non-sentiment
Predicted	sentiment	50	50
	Non-sentiment	150	750

a) What is the accuracy of this classifier?
b) What is the recall, precision and f-score for the sentiment bearing class?
c) What is the recall, precision and f-score for the non-sentiment bearing class?

Exercise 2

J&M, Ch. 4, sec 4.7.1 *Evaluating with more than two classes* shows an example with 3 classes and calculates P, R and F for each class and micro- and macro-averaged P(recision) across the 3 classes. Calculate micro- and macro-averaged R(ecall) and F(-measure) across the 3 classes.

Exercise 3

When applying P, R and F, we should keep in mind that there are two different kinds of scenario. They differ in particular when it comes to micro averaging.

- In the general case, as we see e.g., in information retrieval and span-evaluation for chunking or NER, the total number of predicted items and the total number of gold items may differ.
- In the special case of classification, where a classifier puts each item into exactly one class, there are equally many predicted items as gold items. This was the case in exercise 1 and 2 above.

To see the difference, consider the following. We are evaluating a NER, which considers only three classes, and we count the following numbers.

	True pos.	False pos.	False neg.
Person	720	180	80
Organization	180	20	60
Location	60	0	20

a) Calculate P, R and F for each class
b) Calculate macro- and micro-averaged P, R and F

Part 2: Language models

Consider the small corpus

Corpus 1

This film is funny.
I enjoyed the book.
The film was entertaining.
The book is good.
The game is not bad.
It is not boring.
This is a good book.

We are training an unsmoothed bigram language model (LM) on this corpus. We assume the strings are tokenized by splitting on white space and making punctuation a separate token.

Exercise 1

Which probability will the language model ascribe to the following sequence? Explain how it is calculated.

a) *The film is good.*

Exercise 2

Which problems does this model face if in ascribing a probability to the following sequence?

b) *The film is not good.*

Exercise 3

Modify the model by applying add-one-smoothing and compute the adjusted probabilities for sentence (1) and (2).