Front page

IN4080 Natural Language Processing

Fall 2022

Tuesday, December 13
09:00 AM - 13:00 PM (4 hours)

All questions should be answered!
Each question is assigned a weight which is indicated.
The maximum number of points for the whole set is 100 points.

Permitted materials: None
An on-screen calculator is available.

You may answer in English, Norwegian, Danish or Swedish.

1(a) Naive Bayes

$$\arg\max_{c \in C} P(c \mid \mathrm{f}) = \arg\max_{c \in C} P(c) \prod_{i=1}^{n} P(f_i = v_i \mid c)$$

The formula shows the model for Naive Bayes classification.

- Give a short description of the formula:
  - What is C and c?
  - What is **f**, f_i, v_i and n?
  - What is meant by *argmax*?

Fill in your answer here

Maximum marks: 5

1(b) Assumptions

Which simplifying assumptions are made by the Naive Bayes model? Why can these assumptions result in less accurate classifiers compared to other learning algorithms?

Fill in your answer here

Maximum marks: 5

2(a) HMM tagger

Consider the two sentences

1. February made me shiver.
2. February gave me shiver.

and the two tag sequences

- a) NOUN VERB PRON VERB
- b) NOUN VERB PRON NOUN

It can be argued that the best tag sequence for (1) is (a) and for sentence (2) it is (b). Can a Hidden Markov Model (HMM) tagger be trained to assign tag sequence (a) to sentence (1) and tag sequence (b) to sentence (2)? State reasons for your answer. In case the answer is *yes*, which additional assumptions does the tagger have to fulfill?

Fill in your answer here

Maximum marks: 8

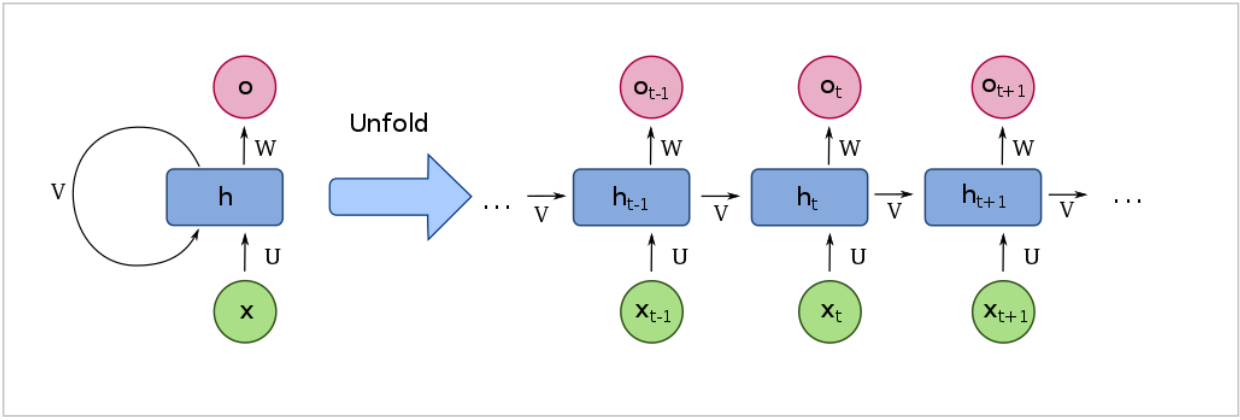
2(b) Discriminative tagger

In mandatory assignment 2, you trained a discriminative Logistic Regression model for POS-tagging. Consider again the example sentences from question (2a) above. Could such a system in principle learn to ascribe tag sequence (a) to sentence (1) and tag sentence (b) to sentence (2)? Explain why or why not. What kind of features would such a tagger need?

Fill in your answer here

Maximum marks: 6

2(c) RNN



The figure, taken from Wikipedia, is meant to illustrate a simple Recurrent Neural Network (RNN). We will use such a network for POS-tagging. Taking sentence (1) from question (2a) as example, answer the following questions:

- To what do x_t and o_t correspond in the case of POS-tagging?
- What is h_t meant to represent?
- How is h_t computed?
- How is o_t computed?

Fill in your answer here

Maximum marks: 8

2(d) Comparison

Shortly compare discriminative Logistic Regression taggers and RNN-taggers from the figure in question (2c). Make the comparison general and do not restrict it to the example sentences above. Do you see any advantages by using the RNN-tagger? Do you see any shortcomings with the simple RNN-tagger from the figure? Can you propose ways to improve the RNN-tagger to overcome these shortcomings?

Fill in your answer here

Maximum marks: 8

3(a) Approaches

In this course we have seen two ways of representing words as vectors based on their distribution in a corpus. The first is based on word-context matrices and the other is called "word embeddings". Describe shortly the main ideas of the two approaches. In particular, compare the two approaches with respect to the form of the vectors and how the vectors are derived.

Fill in your answer here

Maximum marks: 8

3(b) Similarity

Assume the following simplified word vectors.

- girl - (1, 3)
- boy - (2,1)
- princess - (4, 5)

Using cosine for similarity, which word is closer to *girl*? Is it *boy* or *princess*? Explain how you find the answer.

Fill in your answer here

Maximum marks: 6

3(c) Analogies

Word vectors may be used to study semantic properties of words. In particular, one may consider semantic analogies and, for example, ask what is related to *boy* as *princess* is to *girl*. We may symbolize the question as *girl:princess::boy:?*. A way to answer this question is called the *parallelogram method*. Using the example words, explain how the method works.

Fill in your answer here

Maximum marks: 6

4(a) Linguistic foundations

When reviewing the linguistic foundations of human-human dialogues, we mentioned the concept of *alignment*.

1) Explain in 2-3 sentences what this concept of alignment refers to.

2) Why can it be useful to take alignment phenomena into account when designing dialogue systems? Explain in a few sentences.

Fill in your answer here

Maximum marks: 4

4(b) Chatbot models

We have reviewed during the lectures four distinct approaches that can be taken to develop chatbots:

A) rule-based approaches

B) retrieval-based approaches

C) sequence-to-sequence approaches

D) NLU-based approaches

For each of those four approaches, briefly explain:

1) how the approach processes user inputs ("language understanding" step) ;

2) how the output of this language understanding step is represented ;

3) how the approach selects or generates the system response, based on the output from language understanding.

You do not need to provide detailed explanations, a short description in 1-3 sentences is sufficient for each answer. Since there are four approaches and three questions for each, you need to provide 12 short descriptions.

Fill in your answer here

Maximum marks: 12

4(c) Dialogue management

Explain in a few sentences how frame-based dialogue management operates, and how it differs from dialogue management based on finite-state-automata.

Fill in your answer here

Maximum marks: 4

4(d) Reinforcement learning

What is the cumulative expected reward *Q(s,a)* in reinforcement learning, and how can it be computed from the reward function *R(s,a)*, according to Bellman's equation?

Fill in your answer here

Maximum marks: 5

5(a) Debiasing

How can data augmentation be used to reduce the social biases and stereotypes of data-driven NLP models? Explain in a few sentences, and provide an example.

Fill in your answer here

Maximum marks: 5

5(b) Fairness

Imagine that a Norwegian high school has a plan to reduce the proportion of its pupils that fail their mathematics exam. To this end, they develop an automated procedure to decide who should be given extra lessons in mathematics. This automated procedure relies on a machine learning model that predicts the probability of passing the exam based on various information about the pupils and their school achievements through the year.

Before they launch this new procedure, the school would like to ensure their procedure is fair to pupils that have a migration background. They decide to run their machine learning model on a sample of 10 pupils from the previous year, and get the following results:

Student	Migration background?	Prediction from model (pass or fail math exam)	Actual exam outcome
1	No	Pass	Pass
2	Yes	Pass	Fail
3	No	Pass	Pass
4	Yes	Fail	Pass
5	No	Fail	Fail
6	Yes	Pass	Fail
7	No	Fail	Fail
8	Yes	Pass	Pass
9	No	Pass	Pass
10	Yes	Fail	Pass

Analyse the fairness of the model (in relation to the migration background of the pupil) in terms of the "equality of odds" criteria. Show your calculations.

Based on the results you obtained, would you consider the model to be fair to the two groups of pupils? Justify your answer.

Fill in your answer here

Maximum marks: 10