Sektion 1

Instructions

General instructions

This exam consists of two parts:

Part A consists of 5 items (A1–A5), each worth 3 points. These items test your understanding of the basic algorithms that are covered in the course. They require only compact answers, such as a short text, calculation, or diagram.

Part B consists of 3 items (B1–B3), each worth 6 points. These items test your understanding of the more advanced algorithms that are covered in the course. They require detailed and coherent answers with correct terminology.

Note that surplus points in one part do not raise your score in another part.

Grade requirements 729G17

- Grade G: at least 12 points in Part A
- Grade VG: at least 12 points in Part A and at least 14 points in Part B

Grade requirements TDP030

- Grade 3: at least 12 points in Part A
- Grade 4: at least 12 points in Part A and at least 7 points in Part B
- Grade 5: at least 12 points in Part A and at least 14 points in Part B

Good luck!

Instructions on fractions

When instructed to 'answer with a fraction', you should provide a fraction that evaluates to a concrete number. Inside the fraction, you can use other numbers and mathematical operators from the keypad. **You do not need to simplify the fraction or evaluate it yourself.**

For example, all of these fractions evaluate to the number 0.5: $\frac{1}{2}$, $\frac{2}{4}$, $\frac{0.5}{1}$, $\frac{5 \times 10}{10^2}$

Below is a practice question (not graded) to check whether you have understood this instruction. Answer with a fraction.

What is the percentage of days in a normal week whose name starts with the letter S?

1 $\frac{\square}{\square}$

Correct answers:

1 $\frac{2}{7}$ **Saturday, Sunday**

Text classification

Here is a document collection:

document	class
kiwi lion mule	X
mule newt mule	Y
mule mule owl	Y
mule pig	Y

A1.1 Question from the written exam 2021

Use maximum likelihood estimation with add-one smoothing to estimate the word probabilities of a Naive Bayes classifier from the document collection. Assume that the vocabulary consists of all words occurring in the documents. Answer with fractions.

Class	kiwi	lion	mule
X	1	2	3
Y	4	5	6

Correct answers:

1 $\frac{1+1}{3+6}$ 2 $\frac{1+1}{3+6}$ 3 $\frac{1+1}{3+6}$ 4 $\frac{0+1}{8+6}$ 5 $\frac{0+1}{8+6}$ 6 $\frac{5+1}{8+6}$

A1.2 Question from the written exam 2021

Based on the probabilities just estimated, compute the class-specific scores that the Naive Bayes classifier uses to predict the class for the following document:

kiwi lion mule mule mule

Answer with fractions.

Class X	Class Y
1	2

Correct answers:

1 $\frac{1\times(1+1)\times(1+1)\times(1+1)\times(1+1)}{4\times(3+6)\times(3+6)\times(3+6)\times(3+6)}$ 2 $\frac{3\times(0+1)\times(0+1)\times(5+1)\times(5+1)\times(5+1)}{4\times(8+6)\times(8+6)\times(8+6)\times(8+6)\times(8+6)}$

A1.3 Question from the written exam 2021

Here are some class frequencies:

	Class A	Class B	Class C
training data	2958	2465	1972
test data	492	738	615

What is the accuracy of the most frequent class baseline on the test data? Answer with a fraction.

$\frac{\square}{\square}$

Correct answers:

$\frac{492}{492+738+615}$

Language modelling

The WikiText language modeling dataset is a collection of 2 million tokens extracted from verified 'Good' and 'Featured' articles on Wikipedia, comprising a vocabulary of 33,000 unique words. In this dataset we have the following selected counts of unigrams and bigrams:

book	first	the	book first	book the	first book	the book
611	3981	113161	0	1	8	200

A2.1 Question from the written exam 2021

Estimate the following probabilities using maximum likelihood estimation without smoothing. Answer with fractions.

P(book)	P(book | first)
1 $\square$	2 $\square$

Correct answers:

1 $\frac{611}{2000000}$ 2 $\frac{8}{3981}$

A2.2 Question from the written exam 2021

Estimate the following probabilities using maximum likelihood estimation with add-k smoothing, $k = \frac{1}{100}$. Answer with fractions.

P(book)	P(first | book)
1 $\square$	2 $\square$

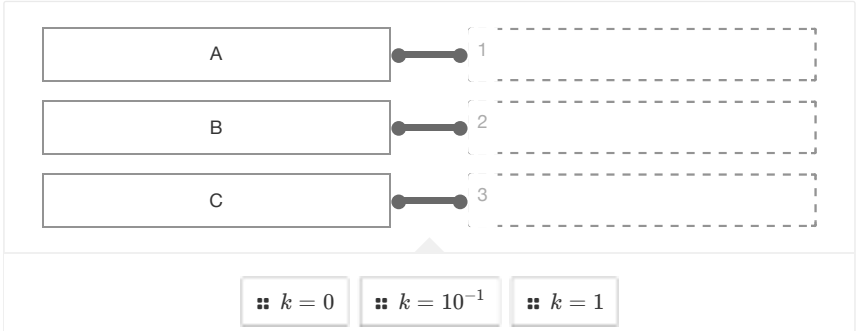
Correct answers:

1 $\frac{611+\frac{1}{100}}{2000000+\frac{1}{100}\times 33000}$ 2 $\frac{0+\frac{1}{100}}{611+\frac{1}{100}\times 33000}$

A2.3 Question from the optional tests 2020

We use maximum likelihood estimation with add-k smoothing to train n-gram models on the WikiText corpus. The following table shows the perplexity of each model on the training data. Which model corresponds to which k-value?

Model	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
A	162	167	351	404	426
B	162	10	3	2	1
C	162	63	106	124	133



Correct answers:

1 $k = 1$ 2 $k = 0$ 3 $k = 10^{-1}$

Part-of-speech tagging

A3.1 Question from the written exam 2021

The evaluation of a part-of-speech tagger produced the confusion matrix shown below. The marked cell gives the number of times the system tagged a word as an adjective (ADJ) whereas the gold standard specified it as a determiner (DET).

	ADJ	DET	NOUN	VERB
ADJ	1475	0	221	31
DET	5	1835	3	0
NOUN	45	5	3887	167
VERB	28	1	387	2135

Set up fractions for the following:

– recall with respect to nouns: 1 $\square$

– precision with respect to determiners: 2 $\square$

Correct answers:

1 $\frac{3887}{45+5+3887+167}$ 2 $\frac{1835}{0+1835+5+1}$

A3.2 Question from the written exam 2021

Training a Hidden Markov Model amounts to estimating two types of probabilities. What is the total number of probability values that you need to estimate when training a model with 17 tags and a vocabulary consisting of 19672 unique words? Answer with a formula that evaluates to a concrete number. Example: 2×3

Do not forget to take into account the beginning-of-sentence and end-of-sentence markers!

The total number of probabilities is 1 [ ].

Correct answers:

1 $(17+1) \times (17+1) + (17 \times 19672)$

Also accepted:
17 x 17 + 17 x 19672

A3.3 Simplified version of a question from the written exam 2019

One difference between a multi-class perceptron tagger and a tagger based on a hidden Markov model (HMM) is in the features that the two models have access to. Which (zero or more) of the following features would you have to choose to provide the multi-class perceptron tagger with the same information that the HMM tagger has access to?

- [x] current word
- [ ] word to the left of the current word
- [ ] word to the right of the current word
- [x] part-of-speech tag of the word to the left of the current word

Syntactic analysis

A4.1 Question from the optional tests 2019

Here are all NP-rules and all VP-rules from a certain probabilistic context-free grammar. Complete the fractions by filling in the missing values.

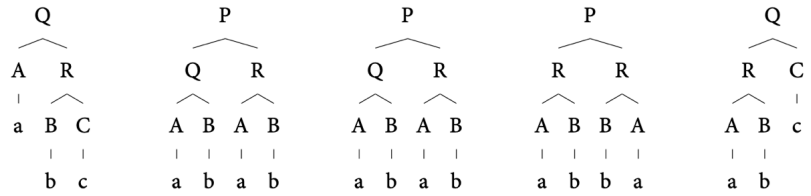
NP → PRP	$\frac{2}{7}$
NP → NP PP	$\frac{1}{7}$
NP → DT NN	$\frac{2}{7}$
NP → NN	1 $\frac{\square}{\square}$
VP → VB NP	2 $\frac{1}{\square}$
VP → VB NP PP	3 $\frac{1}{\square}$

Correct answers:

1 $\frac{2}{7}$ 2 $\frac{1}{2}$ 3 $\frac{1}{2}$

A4.2 Question from the optional tests 2018

Here is a small phrase structure treebank:



Assuming that rule probabilities are estimated using maximum likelihood estimation (no smoothing), what is the probability of the tree in the middle? Answer with a fraction.

$\frac{\square}{\square}$

Correct answers:

$\frac{2\times2\times4}{3\times4\times6}$

A4.3 Simplified version of a question from the optional tests 2020

The following transition sequence creates a dependency tree for a six-word sentence:

SH SH SH SH RA SH SH LA RA LA RA

In this tree, which (zero or more) words are dependents of word 3?

A	Word 1	
B	Word 2	✓
C	Word 3	
D	Word 4	✓
E	Word 5	
F	Word 6	✓

Semantic analysis

A5.1 Question from the written exam 2021

For each of the following pairs of sentences, what is the semantic relation between the emphasized words?

A: The witch gave him a **terrible** curse.

B: Riding the rollercoaster was a **frightening** experience.

Semantic relation: 1

A: She went to **school** in the Netherlands.

B: Our **school** was housed in a 19th-century building in the Old Town.

Semantic relation: 2

A: His team won the **match** 4–3.

B: There was not a single **match** left in the box.

Semantic relation: 3

A: The **greengrocery** is next to the meat market.

B: Many **shops** had to close when the war started.

Semantic relation: 4

Correct answers:

1 synonymy 2 polysemy 3 homonymy 4 hyponymy

A5.2

Here are six synsets from WordNet:

1. eggshell
2. bark
3. natural covering, cover
4. fish scale
5. rock, stone
6. natural object

The following statements about these synsets define a partial WordNet-like hierarchy:

- 1, 2 and 4 are hyponyms of 3
- 3 and 5 are hyponyms of 6

Based on this hierarchy, what is the path-length similarity between **eggshell** and **rock, stone**? Answer with a fraction.

1 $\frac{\square}{\square}$

Correct answers:

1 $\frac{1}{1+3}$

New question (2022)

Concept first introduced in Lecture 5:
https://youtu.be/gdw_7AIGAqU?t=250

Additional example in the interactive session on 2022-02-22:
https://web.microsoftstream.com/video/1cd904c3-92c8-44ab-9740-d8e8f4fc1ed4?st=2940

A5.3

State the distributional hypothesis.

The distributional hypothesis states that we can learn something about the meaning of a word by looking at the other words that this word co-occurs with. Slogan: ‘You shall know a word by the company it keeps.’

0 / 100 Word Limit

New question (2022)

Concept first introduced in Lecture 5:
https://youtu.be/gdw_7AlGAqU?t=479

Additional explanation in the interactive session on 2022-02-22:
https://web.microsoftstream.com/video/1cd904c3-92c8-44ab-9740-d8e8f4fc1ed4?st=3267