Sektion 1

## 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 methods 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 methods 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 729G86**

- Grade E: at least 12 points in Part A
- Grade D: at least 12 points in Part A and at least 4 points in Part B
- Grade C: at least 12 points in Part A and at least 7 points in Part B
- Grade B: at least 12 points in Part A and at least 10 points in Part B
- Grade A: 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 expressions

When instructed to 'answer with an expression', use the keypad to enter a mathematical expression. The expression must evaluate to a number. For example, all of the following expressions evaluate to the number 0.5:

$\frac{1}{2}$, $\frac{2}{4}$, $\frac{4-3}{2\times 1}$, $\frac{5\times 10}{10^2}$

**You do not need to simplify the expression or evaluate it yourself.**

Inside an expression, you can use mathematical operators and other numbers.

- Do not use operators or symbols that are not readily available on the keypad.
- When writing numbers with decimals, use a period, not a comma: write 0.5, not 0,5.
- Do not use thousands separators: write 2000000, not 2 000 000.

When instructed to 'answer with an fraction', answer with an expression in the form of a fraction.

# Test question (not graded)

The year 2023 has 365 days and 52 Mondays. What is the percentage of Mondays among the days of the year 2023? Answer with a fraction.

1

**Correct answers:**

1 $\frac{52}{365}$

Sektion 2

## Text classification

You want to train and evaluate a Naive Bayes classifier that predicts whether a speech held in the Swedish Riksdag was delivered by a left-wing politician or a right-wing politician. As your training and testing data, you use all the speeches held in the Riksdag in two consecutive sessions:

| dataset | # speeches | # left-wing speeches | # right-wing speeches |
|---|---|---|---|
| training | 12637 | 6889 | 5748 |
| testing | 12343 | 6636 | 5707 |

# A1.1

The following table shows the token counts for the training data, broken down by class, along with the token counts for two selected words.

| class | # tokens | token count for 'ekonomi' | token count for 'fred' |
|---|---|---|---|
| left-wing | 1862855 | 285 | 155 |
| right-wing | 1807975 | 266 | 81 |

State the specified word probabilities of your trained classifier. Assume that probabilities are estimated using maximum likelihood estimation without smoothing. Answer with fractions.

| class | class probability | word probability for 'ekonomi' | word probability for 'fred' |
|---|---|---|---|
| left-wing | $\frac{6889}{12637}$ | 1 $\frac{\square}{\square}$ | 2 $\frac{\square}{\square}$ |
| right-wing | $\frac{5748}{12637}$ | 3 $\frac{\square}{\square}$ | 4 $\frac{\square}{\square}$ |

**Correct answers:**

1 $\frac{285}{1862855}$ 2 $\frac{155}{1862855}$ 3 $\frac{266}{1807975}$ 4 $\frac{81}{1807975}$

# A1.2

Suppose that your classifier also contains the following word probabilities:

| class | word probability for ‘linköping’ | word probability for ‘student’ |
|---|---|---|
| left-wing | $\frac{924}{78239910}$ | $\frac{924}{78239910}$ |
| right-wing | $\frac{966}{75934950}$ | $\frac{168}{75934950}$ |

What are the class-specific scores that your Naive Bayes classifier uses to predict the class for the following document:

**linköping student student**

Answer with expressions.

| left-wing | right-wing |
|---|---|
| 1 | 2 |

### Correct answers:

1 $\frac{6889}{12637} \times \frac{924}{78239910} \times \left(\frac{924}{78239910}\right)^2$ 2 $\frac{5748}{12637} \times \frac{966}{75934950} \times \left(\frac{168}{75934950}\right)^2$

# A1.3

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

### Correct answers:

$\frac{6636}{12343}$

Sektion 3

## Language modelling

We consider a language modelling dataset collected from Arthur Conan Doyle’s short stories about Sherlock Holmes. The dataset contains 384046 tokens and comprises a vocabulary of 15339 unique words. We have the following selected counts of unigrams and bigrams:

| dear | sherlock | holmes | dear holmes | holmes dear | sherlock holmes | holmes sherlock |
|---|---|---|---|---|---|---|
| 191 | 210 | 1492 | 7 | 0 | 195 | 0 |

# A2.1

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

| P(sherlock) | P(sherlock \| holmes) |
|---|---|
| 1 | 2 |

**Correct answers:**

1 $\frac{210}{384046}$ 2 $\frac{0}{1492}$

# A2.2

Estimate the following probabilities using maximum likelihood estimation with add-$k$ smoothing, $k = \frac{1}{4}$. Answer with fractions.

| P(sherlock) | P(holmes \| sherlock) |
|---|---|
| 1 | 2 |

**Correct answers:**

1 $\frac{210+\frac{1}{4}}{384046+\frac{1}{4}\times 15339}$ 2 $\frac{195+\frac{1}{4}}{210+\frac{1}{4}\times 15339}$

# A2.3

Suppose you have trained a unigram model with the following probabilities:

$P(\text{foo}) = \frac{1}{2} \quad P(\text{bar}) = \frac{1}{4} \quad P(\text{baz}) = \frac{1}{8}$

What is the entropy of your unigram model on the following test sentence?

**foo bar bar baz**

Answer with a single number.

To help you, here are some base-2 logarithms:

$\log_2 \frac{1}{1} = 0 \quad \log_2 \frac{1}{2} = -1 \quad \log_2 \frac{1}{4} = -2 \quad \log_2 \frac{1}{8} = -3 \quad \log_2 \frac{1}{16} = -4 \quad \log_2 \frac{1}{32} = -5 \quad \log_2 \frac{1}{64} = -6 \quad \log_2 \frac{1}{128} = -7 \quad \log_2 \frac{1}{256} = -8$

1

**Correct answers:**

1 2

Sektion 4

# A3.1

The evaluation of a part-of-speech tagger produced the confusion matrix shown below. The marked cell contains the number of times the tagger 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 |

What is the F1-score with respect to nouns? Answer with a fraction.

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

### Correct answers:

1 $$\frac{2\times\frac{3887}{221+3+3887+387}\times\frac{3887}{45+5+3887+167}}{\frac{3887}{221+3+3887+387}+\frac{3887}{45+5+3887+167}}$$

# A3.2

The following matrices specify (parts of) a hidden Markov model. The marked cell specifies the probability for the transition from BOS to AB.

**Transition probabilities**

| | AB | PN | PP | VB | EOS |
|---|---|---|---|---|---|
| **BOS** | $\frac{1}{11}$ | $\frac{1}{10}$ | $\frac{1}{12}$ | $\frac{1}{11}$ | $\frac{1}{25}$ |
| **AB** | $\frac{1}{11}$ | $\frac{1}{11}$ | $\frac{1}{11}$ | $\frac{1}{10}$ | $\frac{1}{14}$ |
| **PN** | $\frac{1}{11}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{10}$ | $\frac{1}{16}$ |
| **PP** | $\frac{1}{13}$ | $\frac{1}{11}$ | $\frac{1}{12}$ | $\frac{1}{14}$ | $\frac{1}{18}$ |
| **VB** | $\frac{1}{11}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{13}$ | $\frac{1}{15}$ |

**Emission probabilities**

| | she | got | up |
|---|---|---|---|
| **AB** | $\frac{1}{25}$ | $\frac{1}{25}$ | $\frac{1}{14}$ |
| **PN** | $\frac{1}{13}$ | $\frac{1}{25}$ | $\frac{1}{25}$ |
| **PP** | $\frac{1}{25}$ | $\frac{1}{25}$ | $\frac{1}{13}$ |
| **VB** | $\frac{1}{25}$ | $\frac{1}{14}$ | $\frac{1}{19}$ |

Under this model, what is the probability of the following tagged sentence? Answer with an expression.

| she | got | up |
|---|---|---|
| PN | VB | AB |

1

Correct answers:

1 $\frac{1}{10} \times \frac{1}{13} \times \frac{1}{10} \times \frac{1}{14} \times \frac{1}{11} \times \frac{1}{14} \times \frac{1}{14}$

# A3.3

Consider a part-of-speech tagger based on the multi-class perceptron with a feature window containing the following features:

1. tag of the previous word
2. current word
3. next word

The following table shows the values of these features when tagging the sentence

**I want to live in peace**

Complete the missing values. If a feature value is not defined, use the ‘undefined’ card.

| feature 1 | feature 2 | feature 3 | predicted tag |
|---|---|---|---|
| 1 | 2 | 3 | PRON |
| 4 | 5 | 6 | VERB |
| 7 | 8 | 9 | PART |
| 10 | 11 | 12 | VERB |
| 13 | 14 | 15 | ADP |
| 16 | 17 | 18 | NOUN |

I | want | to | live | in | peace | PRON | VERB | PART | ADP | NOUN | undefined

**Correct answers:**

1 undefined 2 I 3 want 4 PRON 5 want 6 to 7 VERB

8 to 9 live 10 PART 11 live 12 in 13 VERB 14 in 15 peace

16 ADP 17 peace 18 undefined

**Sektion 5**

# A4.1

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

| | |
|---|---|
| NP → DT NN | 1 $\frac{\square}{\square}$ |

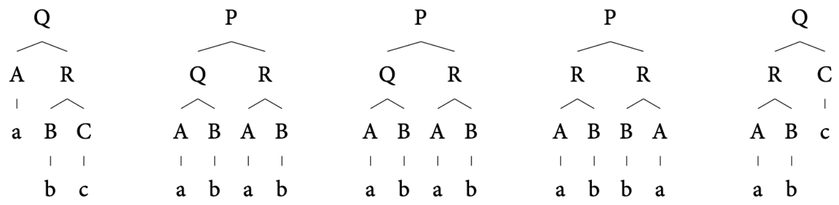
| | |
|---|---|
| NP → NN | $\frac{1}{3}$ |
| NP → NP PP | $\frac{4}{9}$ |
| VP → VB NP | 2 |
| VP → VB NP PP | $\frac{8}{9}$ |

**Correct answers:**

1 $\frac{2}{9}$ 2 $\frac{1}{9}$

# A4.2

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 fourth tree from the left? Answer with an expression.

**Correct answers:**

$\frac{1}{3} \times \frac{4}{6} \times \frac{1}{6}$

# A4.3

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

**SH SH SH LA SH SH SH LA RA LA RA**

State an alternative transition sequence for the same dependency tree.

1 2 3 4 5 6 7
8 9 10 11

SH LA RA

**Correct answers:**

1 SH 2 SH 3 SH 4 LA 5 SH 6 LA 7 SH 8 SH 9 LA
10 RA 11 RA

Sektion 6

# A5.1

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

A: The lake was **calm**, creating a peaceful atmosphere.

B: During the exam, the students were **quiet**, trying to focus on their answers.

**Semantic relation:** 1

A: Tracy is always **cheerful**.

B: John's recent breakup with Tracy has left him feeling **gloomy** and depressed.

**Semantic relation:** 2

A: She regretted squeezing her feet into **heels** the morning after the party.

B: It is important to wear the right kind of **footwear** when hiking in the mountains.

**Semantic relation:** 3

A: He deposited his paycheck at the **bank**.

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

**Semantic relation:** 4

Correct answers:

1 synonymy 2 antonymy 3 hyponymy 4 polysemy

# A5.2

Here are six synsets from WordNet:

1. preschool
2. educational institution
3. central bank
4. university
5. institution, establishment
6. financial institution

Arrange these synsets into a hierarchy as in WordNet. Based on this hierarchy, what is the path-length similarity between **preschool** and **central bank**? Answer with a fraction.

1

Correct answers:

1 $\frac{1}{1+4}$

# A5.3

We read off word vectors from the following co-occurrence matrix (target words correspond to rows, context words correspond to columns):

| | **caws** | **dafad** |
|---|---|---|
| **cheese** | 6 | 2 |
| **sheep** | 0 | 4 |
| **goat** | 1 | 6 |
| **bread** | 5 | 0 |

Order the four target words in increasing degree of semantic similarity to the word **bread** (least similar at the top, most similar at the bottom). Assume that semantic similarity is measured in terms of cosine similarity between word vectors.

≡ sheep

1 **Correct answer:** sheep

≡ goat

2 **Correct answer:** goat

≡ cheese

3 **Correct answer:** cheese

≡ bread

4 **Correct answer:** bread

Sektion 7

## Edit distance

**You only need to work on this item if you are aiming for a higher grade than Pass.**

# B1.1

Define the concept of the Levenshtein distance between two words. The definition should be detailed and understandable even to readers who have not taken this course.

0 / 500 Word Limit

# B1.2

Compute the Levenshtein distance between the two words **fiction** and **fashion** using the Wagner–Fischer algorithm. Complete the matrix and state the distance.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| o | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| i | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| t | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| c | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
| i | 41 | 42 | 43 | 44 | 45 | 46 | 47 |
| f | 49 | 50 | 51 | 52 | 53 | 54 | 55 |
| # | 57 | 58 | 59 | 60 | 61 | 62 | 63 |
| | # | f | a | s | h | i | o |

The Levenshtein distance is: 65

**Correct answers:**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 2 | 6 | 3 | 6 | 4 | 6 | 5 | 6 | 6 | 5 | 7 | 4 | 8 | 3 | 9 | 6 | 10 | 5 |
| 11 | 5 | 12 | 5 | 13 | 5 | 14 | 4 | 15 | 3 | 16 | 4 | 17 | 5 | 18 | 4 | 19 | 4 | 20 | 4 |
| 21 | 4 | 22 | 3 | 23 | 4 | 24 | 5 | 25 | 4 | 26 | 3 | 27 | 3 | 28 | 3 | 29 | 3 | 30 | 4 |
| 31 | 5 | 32 | 5 | 33 | 3 | 34 | 2 | 35 | 2 | 36 | 2 | 37 | 3 | 38 | 4 | 39 | 4 | 40 | 5 |
| 41 | 2 | 42 | 1 | 43 | 1 | 44 | 2 | 45 | 3 | 46 | 3 | 47 | 4 | 48 | 5 | 49 | 1 | 50 | 0 |
| 51 | 1 | 52 | 2 | 53 | 3 | 54 | 4 | 55 | 5 | 56 | 6 | 57 | 0 | 58 | 1 | 59 | 2 | 60 | 3 |
| 61 | 4 | 62 | 5 | 63 | 6 | 64 | 7 | 65 | 3 | | | | | | | | | | |

# B1.3

A **subsequence** of a given word is a sequence of characters from that word. The characters must come in the same order as in the original word, but they do not need to be consecutive. For example, the following sequences are subsequences of the word *fiction*: *fi*, *on*, *fion*. On the other hand, these are not subsequences of *fiction*: *ac*, *if*, *noi*. A **longest common subsequence (LCS)** is the longest subsequence common to two words. For example, the LCS of the words *fiction* and *instruction* is *iction*.

Edit distance and LCS are closely related. To see this, note every word can be transformed into every other word via their LCS:

1. Delete all characters in the first word that are not part of the LCS.
2. Insert all characters of the second word that are not parts of the LCS.

For example (LCS in bold):

*f**iction** → **iction** → **i**nstru**ction***

Based on this observation, explain how a simplified version of the Wagner–Fisher algorithm can be used to compute the <u>length</u> of the LCS of two words.

0 / 500 Word Limit

Sektion 8

## Viterbi algorithm

**You only need to work on this item if you are aiming for a higher grade than Pass.**

# B2.1

Here is a Hidden Markov model specified in terms of costs (negative log probabilities).

**Transition costs**

The topmost, leftmost number is the transition cost from BOS to ADP.

| | ADP | ADV | PRON | VERB | EOS |
|---|---|---|---|---|---|
| **BOS** | 14 | 14 | 12 | 14 | 27 |
| **ADP** | 15 | 15 | 13 | 16 | 21 |
| **ADV** | 13 | 13 | 13 | 13 | 17 |
| **PRON** | 14 | 14 | 15 | 12 | 18 |
| **VERB** | 13 | 14 | 13 | 15 | 17 |

**Emission costs**

| | she | got | up |
|---|---|---|---|
| **ADP** | 28 | 28 | 15 |
| **ADV** | 27 | 27 | 16 |
| **PRON** | 16 | 28 | 28 |
| **VERB** | 28 | 16 | 21 |

When using the Viterbi algorithm to calculate the least expensive (most probable) tag sequence for the sentence 'she got up', one gets the matrix below. Fill in the missing values.

| | | she | got | up | |
|---|---|---|---|---|---|
| BOS | 0 | | | | |
| ADP | | 42 | 70 | 1 | |
| ADV | | 41 | 69 | 2 | |
| PRON | | 28 | 71 | 3 | |
| VERB | | 42 | 56 | 4 | |
| EOS | | | | | 5 |

Correct answers:

1 84 2 86 3 97 4 92 5 103

# B2.2

Let $m$ and $n$ denote the number of tags in the HMM and the number of words in the input sentence, respectively. The following statements hold:

1. The number of tagged sentences grows exponentially with the sentence length. More specifically, that number is $m^n$.
2. In spite of this, the time that the Viterbi algorithm needs to find the best tag sequence is polynomial. More specifically, that number is in $O(m^2n)$

Explain what these statements mean and argue why they hold.

0 / 500 Word Limit

# B2.3

Let $m$ and $n$ be as defined as above. The memory required by the Viterbi algorithm is in $O(mn)$. However, when one is only interested in the **cost** of the least expensive tag sequence, not in the sequence itself, the memory required by the Viterbi algorithm is in $O(m)$. Explain this statement and argue why it holds. Why does the statement **not** hold if one wants to reconstruct the actual tag sequence?

0 / 500 Word Limit

Sektion 9

## Parsing probabilistic context-free grammars

**You only need to work on this item if you are aiming for a higher grade than Pass.**

Here is a fragment of a probabilistic context-free grammar (PCFG). It is specified here in terms of costs (negative log probabilities) instead of regular probabilities:

| rule | cost |
|---|---|
| S → NP VP | 1 |
| NP → Det N | 5 |
| VP → V NP | 7 |
| V → includes | 13 |
| Det → a | 4 |
| Det → the | 4 |
| N → flight | 17 |
| N → meal | 20 |

We will use this grammar to parse the following sentence:

**the flight includes a meal**

# B3.1

State the problem solved by the probabilistic extension of the CKY algorithm when applied to this grammar. Your statement should be detailed and understandable even to readers who have not taken this course.

0 / 500 Word Limit

# B3.2

The CKY algorithm assumes the input PCFG to be in **Chomsky normal form**. Explain what this restriction means, and provide an example of a grammar that does not satisfy this restriction.

0 / 500 Word Limit

# B3.3

Provide the full probabilistic CKY chart for the example sentence. Note that instead of multiplying probabilities, you should now add the corresponding costs. The other operations of the CKY algorithm remain unchanged.

| **the** | **flight** | **includes** | **a** | **meal** |
|---|---|---|---|---|
| [0, 1] Det 1 | [0, 2] NP 2 | [0, 3] | [0, 4] | [0, 5] S 3 |
| | [1, 2] N 4 | [1, 3] | [1, 4] | [1, 5] |
| | | [2, 3] V 5 | [2, 4] | [2, 5] VP 6 |
| | | | [3, 4] Det 7 | [3, 5] NP 8 |
| | | | | [4, 5] N 9 |

Correct answers:

1 4 2 26 3 76 4 17 5 13 6 49 7 4 8 29 9 20