— IN2110 —
Methods in Language Technology

Summing up
Exam preparations

Stephan Oepen & Erik Velldal

Language Technology Group (LTG)

May 21, 2019

Topics for today

- High-level summary
- Practical details regarding the final exam
- Sample exam questions (though not a sample exam)

Reading List

- Both the lecture notes (slides) and the background reading specified in the lecture schedule (at the course page) are obligatory reading.
- We also expect that you have looked at the provided model solutions for the exercises.

Final Written Examination

When / where:

- 14 June at 09:00 (4 hours)
- Sal 3D Silurveien 2 (double-check with StudentWeb)
- Digital exam, Inspera

Final Written Examination

When / where:

- 14 June at 09:00 (4 hours)
- Sal 3D Silurveien 2 (double-check with StudentWeb)
- Digital exam, Inspera

The exam

- When writing your answers, remember...
  - ~~Less~~ more is more! (As long as it's relevant.)
  - Aim for high recall *and* precision.
  - Don't just list keywords; spell out what you think.
  - If you see an opportunity to show off terminology, seize it.
  - Each question will have points attached (summing to 100) to give you an idea of how they will be weighted in the grading.

Inventory

Main areas

- Vector space models; representing words and documents
- Classification
- Clustering
- Sequence modeling
- Syntax and parsing
- Evaluation methodology (metrics and data splits)

Progression

- Representation
- From geometric to probabilistic models
- From 'point-wise' to sequential and finally hierarchical modeling $\longrightarrow$

Problems we have dealt with

- How to model similarity relations between point-wise observations, and how to represent and predict group membership.
- E.g. vector space models and classification over words and documents.

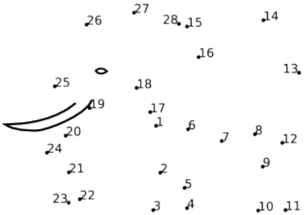
Problems we have dealt with

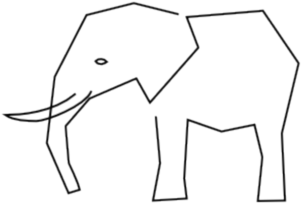
- How to model similarity relations between point-wise observations, and how to represent and predict group membership.
- E.g. vector space models and classification over words and documents.
- Sequences
  - Probabilities over strings: Markov chains and n-gram models: Linear and surface oriented.
  - Sequence classification: HMMs and CRF add one layer of abstraction; class labels as hidden variables. But still only linear.

Problems we have dealt with

- How to model similarity relations between point-wise observations, and how to represent and predict group membership.
- E.g. vector space models and classification over words and documents.
- Sequences
  - Probabilities over strings: Markov chains and n-gram models: Linear and surface oriented.
  - Sequence classification: HMMs and CRF add one layer of abstraction; class labels as hidden variables. But still only linear.
- Grammar; adds hierarchical structure
  - Shift focus from 'sequences' to 'sentences'.
  - Identifying underlying structure using formal rules.
  - Phrase structure and dependency grammars
    - Declarative aspect: formal grammar.
    - Procedural aspect: parsing strategy.
  - Learn probability distributions over trees or transition sequences.

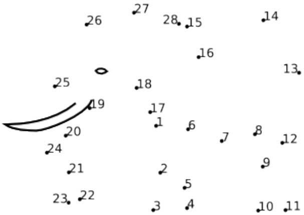
Connecting the dots. . .

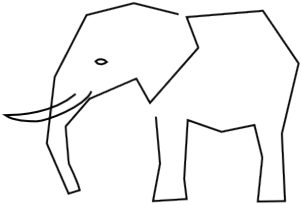




What have we been doing?

Connecting the dots. . .

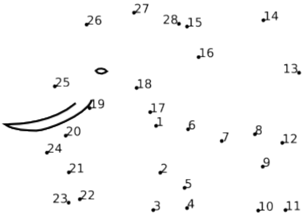


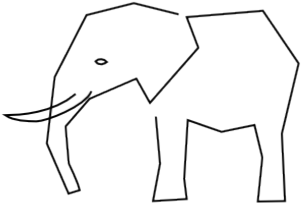


What have we been doing?

- Data-driven learning

Connecting the dots. . .

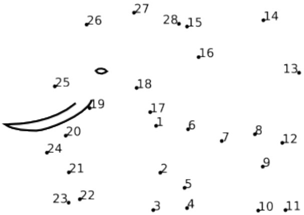


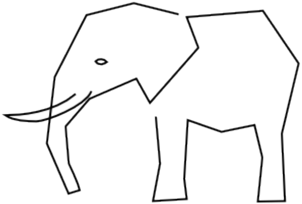


What have we been doing?

- Data-driven learning
- by counting observations

Connecting the dots. . .

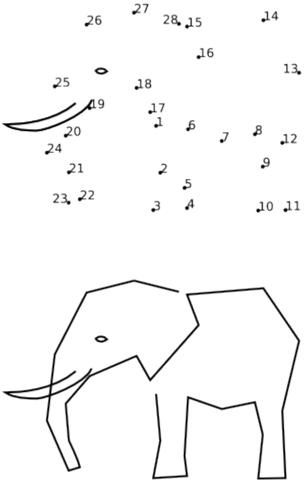




What have we been doing?

- Data-driven learning
- by counting observations
- in context;

Connecting the dots. . .



What have we been doing?

- Data-driven learning
- by counting observations
- in context;
  - context words in vector space models; bag-of-words, etc.
  - previous n-1 words in n-gram models
  - previous n-1 states in HMMs
  - local sub-trees in PCFGs
  - features of configurations in dependency parsing
  - $++$

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b
- all survivors qualified for the final exam ...

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b
- all survivors qualified for the final exam ...
- ... some with a larger margin than others

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b
- all survivors qualified for the final exam ...
- ... some with a larger margin than others
- Three of you stand out in terms of points throughout the term

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b
- all survivors qualified for the final exam ...
- ... some with a larger margin than others
- Three of you stand out in terms of points throughout the term
- A total of 39 points (of 40), we think, is no small accomplishment

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b
- all survivors qualified for the final exam ...
- ... some with a larger margin than others
- Three of you stand out in terms of points throughout the term
- A total of 39 points (of 40), we think, is no small accomplishment
- And the ‘winners’ are:

Finally, Some Statistics

- 62 submitted for oblig 1a, 44 for oblig 2b
- all survivors qualified for the final exam ...
- ... some with a larger margin than others
- Three of you stand out in terms of points throughout the term
- A total of 39 points (of 40), we think, is no small accomplishment
- And the 'winners' are:
  - Magnus Holm (magho)
  - Yauhen Khutarniuk (yauhenk)
  - Kristian Løseth (krislos)
- Great work — Congratulations!

After IN2110

- Please remember to participate in the course evaluation hosted by FUI.
  - Even if this means just repeating the comments you already gave for the midterm evaluation.
  - While the midterm evaluation was only read by us, the FUI course evaluation is distributed department-wide.

After IN2110

- Please remember to participate in the course evaluation hosted by FUI.
  - Even if this means just repeating the comments you already gave for the midterm evaluation.
  - While the midterm evaluation was only read by us, the FUI course evaluation is distributed department-wide.
- Some other courses of potential interest:
- IN3120/IN4120 – Search technology
  - Fall 2019
  - Also based on the book by Manning, Raghavan, & Schütze (2008), `Introduction to Information Retrieval`
- IN3050/IN4050 – Introduction to AI and machine learning (spring 2020)
- We also hope to see many of you in IN2040 – Functional programming in the fall!

Exam prep

- The following are questions that are representative of what you might get at the exam,
- though not a sample exam.

Vector representations

- What is the distributional hypothesis?

Vector representations

- What is the distributional hypothesis?
- Explain what we mean by a *bag-of-words representation* of text (e.g. sentences and documents). Discuss some of the weaknesses of this representation.

Vector representations

- What is the distributional hypothesis?
- Explain what we mean by a *bag-of-words representation* of text (e.g. sentences and documents). Discuss some of the weaknesses of this representation.
- Discuss some high-level differences and similarities between traditional 'count-based' word vectors and 'prediction-based' representations like those computed using more recent approaches like word2vec.

Vector representations

- What is the distributional hypothesis?
- Explain what we mean by a *bag-of-words representation* of text (e.g. sentences and documents). Discuss some of the weaknesses of this representation.
- Discuss some high-level differences and similarities between traditional ‘count-based’ word vectors and ‘prediction-based’ representations like those computed using more recent approaches like word2vec.
- Discuss similarities and differences between Euclidean distance and the cosine measure.

- Explain the difference between supervised and unsupervised learning. For both approaches, mention examples of models that we've touched on throughout the course.

Classification, clustering and ML at large

- Explain the difference between supervised and unsupervised learning. For both approaches, mention examples of models that we've touched on throughout the course.
- What are the differences and similarities between K-means and Rocchio?

Classification, clustering and ML at large

- Explain the difference between supervised and unsupervised learning. For both approaches, mention examples of models that we've touched on throughout the course.
- What are the differences and similarities between K-means and Rocchio?
- In the context of model evaluation, briefly describe what 'micro-averaging' and 'macro-averaging' means, including their differences.

Sequences and structured prediction

- Abstractly, NER is a sequence segmentation task, but in practise it is still approached as a word-by-word sequence-labeling task. How can we represent the labels to facilitate this?
- What is the *Markov assumption*, both generally and in the specific context of HMMs for PoS-tagging?

Sequences and structured prediction

- Abstractly, NER is a sequence segmentation task, but in practise it is still approached as a word-by-word sequence-labeling task. How can we represent the labels to facilitate this?
- What is the *Markov assumption*, both generally and in the specific context of HMMs for PoS-tagging?
- What is dynamic programming? Name some of the dynamic programming algorithms we have encountered in the course and what they are used for.

Exercise (1): Natural Language Ambiguity

Assume the following 'toy' grammar of English:

$$\begin{aligned} S &\rightarrow NP \\ NP &\rightarrow Det\ N \\ N &\rightarrow N\ N \end{aligned}$$

$$\begin{aligned} Det &\rightarrow \textit{the} \\ N &\rightarrow \textit{kitchen} \mid \textit{gold} \mid \textit{towel} \mid \textit{rack} \end{aligned}$$

Exercise (1): Natural Language Ambiguity

Assume the following 'toy' grammar of English:

$$S \rightarrow NP$$
$$NP \rightarrow Det\ N$$
$$N \rightarrow N\ N$$

$$Det \rightarrow \textit{the}$$
$$N \rightarrow \textit{kitchen} \mid \textit{gold} \mid \textit{towel} \mid \textit{rack}$$

(1) How many different syntactic analyses, if any, does the grammar assign to the following strings?

(a) *the kitchen towel rack*
(b) *the kitchen gold towel rack*

Exercise (2): CKY Parsing

Assume the following grammar and CKY parse table:

S → NP VP
VP → V NP
VP → VP PP
NP → NP VP
PP → P NP

	1	2	3	4	5
0	NP		S		S
1		V	VP		VP
2			NP		NP
3				P	PP
4					NP

Exercise (2): CKY Parsing

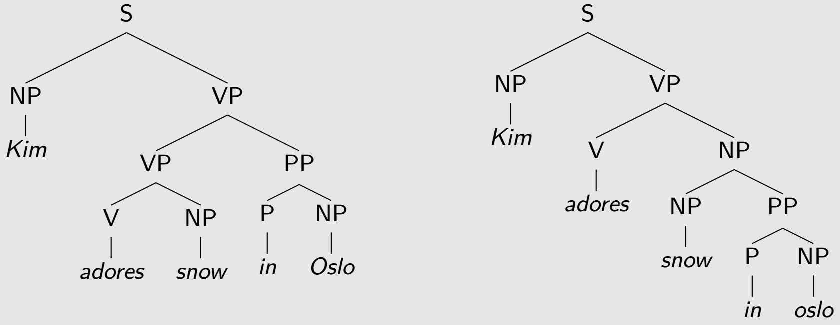
Assume the following grammar and CKY parse table:

$S \rightarrow NP\ VP$
$VP \rightarrow V\ NP$
$VP \rightarrow VP\ PP$
$NP \rightarrow NP\ VP$
$PP \rightarrow P\ NP$

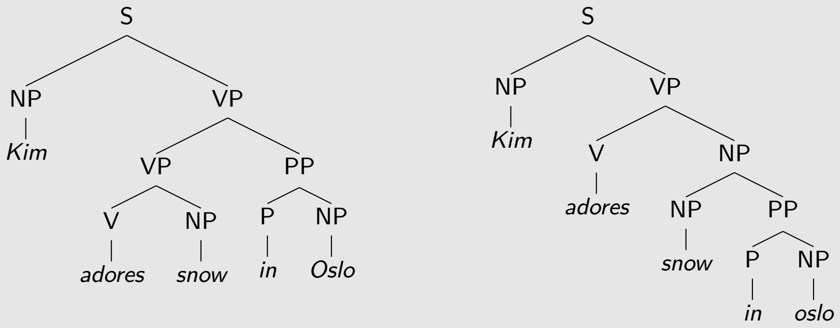
	1	2	3	4	5
0	NP		S		S
1		V	VP		VP
2			NP		NP
3				P	PP
4					NP

(2) Which pair(s) of 'input' cells and which production(s) give rise to the derivation of category S in 'target' cell $\langle 0, 5 \rangle$?

Exercise (4): Dependency Syntaxx

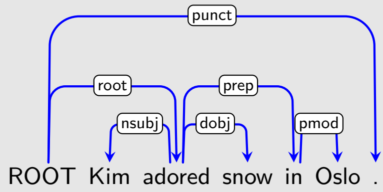


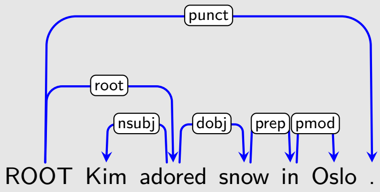
Exercise (4): Dependency Syntaxx



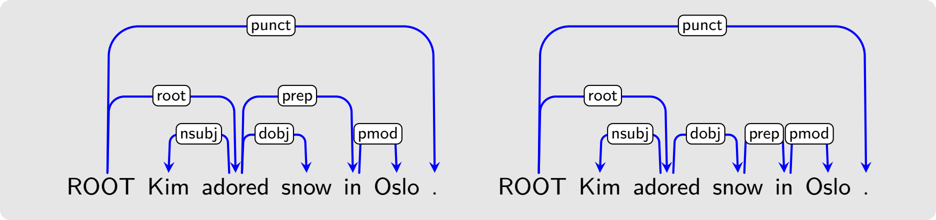
(4) Draw the dependency trees for the two readings.
Where does the attachment ambiguity manifest itself?

Exercise (5): Dependency Evaluation





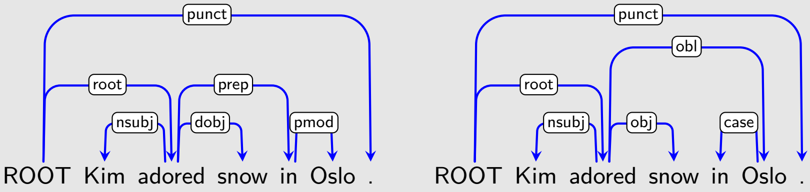
Exercise (5): Dependency Evaluation



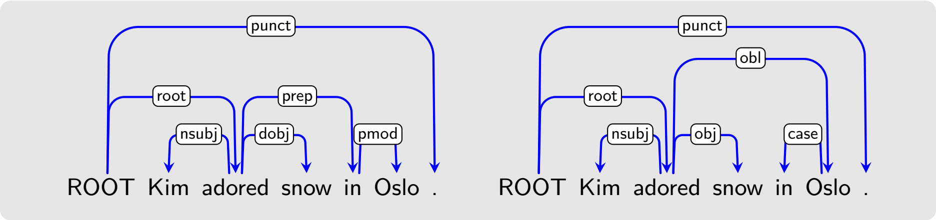
(5) What are the LAS and UAS scores for the two trees?

Gold standard on the left, system prediction on the right.

Exercise (6): More Dependency Evaluation

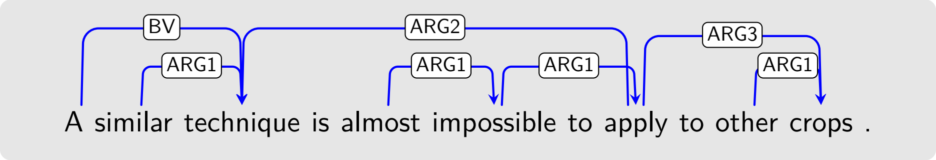


Exercise (6): More Dependency Evaluation

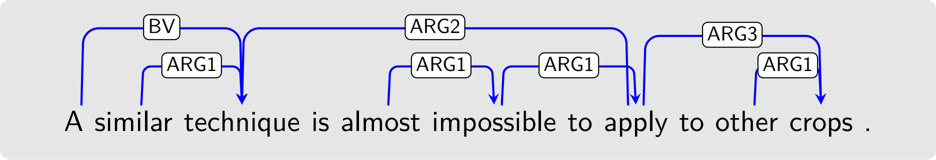


(6) What are the LAS and UAS scores for the two trees?

Exercise (7): Properties of Dependency Graphs



Exercise (7): Properties of Dependency Graphs



(7) Which of the following formal properties hold for this dependency graph: (a) Connectedness, (b) Acyclicity, (c) Single-Headedness, and (d) Projectivity? Explain your answers.