Introduction to Machine Learning

Class #12, May 2 2023

Barbora Hladká hladka@ufal.mff.cuni.cz

Data Analytics for Students of Social Studies and Humanities https://ufal.mff.cuni.cz/courses/npfl134

Task of Word Sense Disambiguation

Assign the correct sense of line in the following sentences.

Consider these senses: `phone`, `formation`, `text`, `cord`, `product`, `division`

1. I've got Inspector Jackson on the line for you.
2. Outside, a line of customers waited to get in.
3. He quoted a few lines from Shakespeare.
4. He didn't catch many fish, but it hardly mattered.
   With his line out, he sat for hours staring at the Atlantic.
5. The company has just launched a new line of small, low-priced computers.
6. Draw a line that passes through the points P and Q.
7. This has been a very popular new line.

Task of Word Sense Disambiguation

Assign the correct sense of line in the following sentences.

Consider these senses: `phone`, `formation`, `text`, `cord`, `product`, `division`

1. I've got Inspector Jackson on the line for you. `phone`
2. Outside, a line of customers waited to get in. `formation`
3. He quoted a few lines from Shakespeare. `text`
4. He didn't catch many fish, but it hardly mattered.
   With his line out, he sat for hours staring at the Atlantic. `cord`
5. The company has just launched a new line of small, low-priced computers. `product`
6. Draw a line that passes through the points P and Q. `division`
7. This has been a very popular new line. `product? formation?`

Task of Word Sense Disambiguation

- What knowledge did you use to assign the senses?
- What were the keys for the correct decision?
- Which sentences were easy to recognize the correct sense and which were the most difficult for you?

Machine Learning

deals with teaching computers to learn from data presented as examples.

- We – human beings – do word sense disambiguation easily using the context in the sentence and our knowledge of the world.
- We want computers to master it as well.
- Could you implement the disambiguation procedure?
- How well would your code perform the task? Does it make errors?

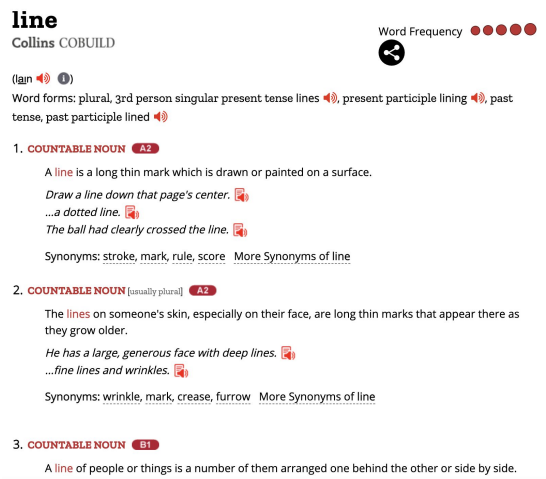
Let's prepare examples and guide computers to learn from them.

That is Machine learning!

A non-technical view on ML

Teach kids to do word sense disambiguation of line.

1. You are a teacher.
2. Lend English dictionary in a library.
3. Discuss with kids the meanings of line listed in the dictionary.
   Focus on the context of line in example sentences.
4. Prepare a quiz to test the kids.
   Select new sentences that kids have NOT seen in the dictionary.
5. Evaluate kids' answers.



line
Collins COBUILD

Word Frequency

(lain)

Word forms: plural, 3rd person singular present tense lines, present participle lining, past tense, past participle lined

1. COUNTABLE NOUN A2

A line is a long thin mark which is drawn or painted on a surface.

Draw a line down that page's center.
...a dotted line.
The ball had clearly crossed the line.

Synonyms: stroke, mark, rule, score More Synonyms of line

2. COUNTABLE NOUN [usually plural] A2

The lines on someone's skin, especially on their face, are long thin marks that appear there as they grow older.

He has a large, generous face with deep lines.
...fine lines and wrinkles.

Synonyms: wrinkle, mark, crease, furrow More Synonyms of line

3. COUNTABLE NOUN B1

A line of people or things is a number of them arranged one behind the other or side by side.

A non-technical view on ML

Will all the kids get "A" grade?

- May be, and may be not. Why?
  They have not seen all possible sentences with line during their learning.

A non-technical view on ML

How to get better grades?

- Get more example sentences, e.g. lend more English dictionaries.
- Focus on other specific characteristics, e.g. part of speech classes that occur in the context of line.

Machine learning

Teach computers to learn from examples in five essential steps

1. Formulating the task ("Assign the correct sense of line in a sentence.")
2. Getting examples (= gold data, e.g. manual labelling)
   Splitting them into training and test examples.
3. Learning from training data using ML methods (= algorithms)
4. Testing the learned knowledge (= model) on test data
5. Evaluation

Formal definition of Machine learning (Tom Mitchell, 1997)

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

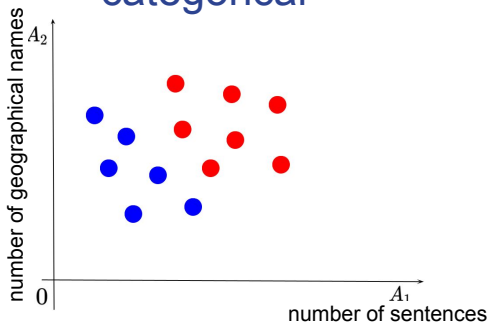
I.e., a machine learning problem is defined by a program learning from experience E with respect to a task T while its success is being measured by the performance (measure) P which should improve at performing task T with experience E.

Supervised machine learning

- is building a model from examples for predicting an output for an input. We, teachers (supervisors), prepare examples.
- input = vector representing a real world object
- output

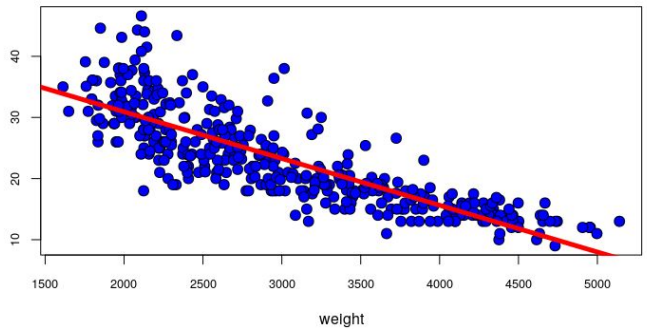
categorical

classification



numerical

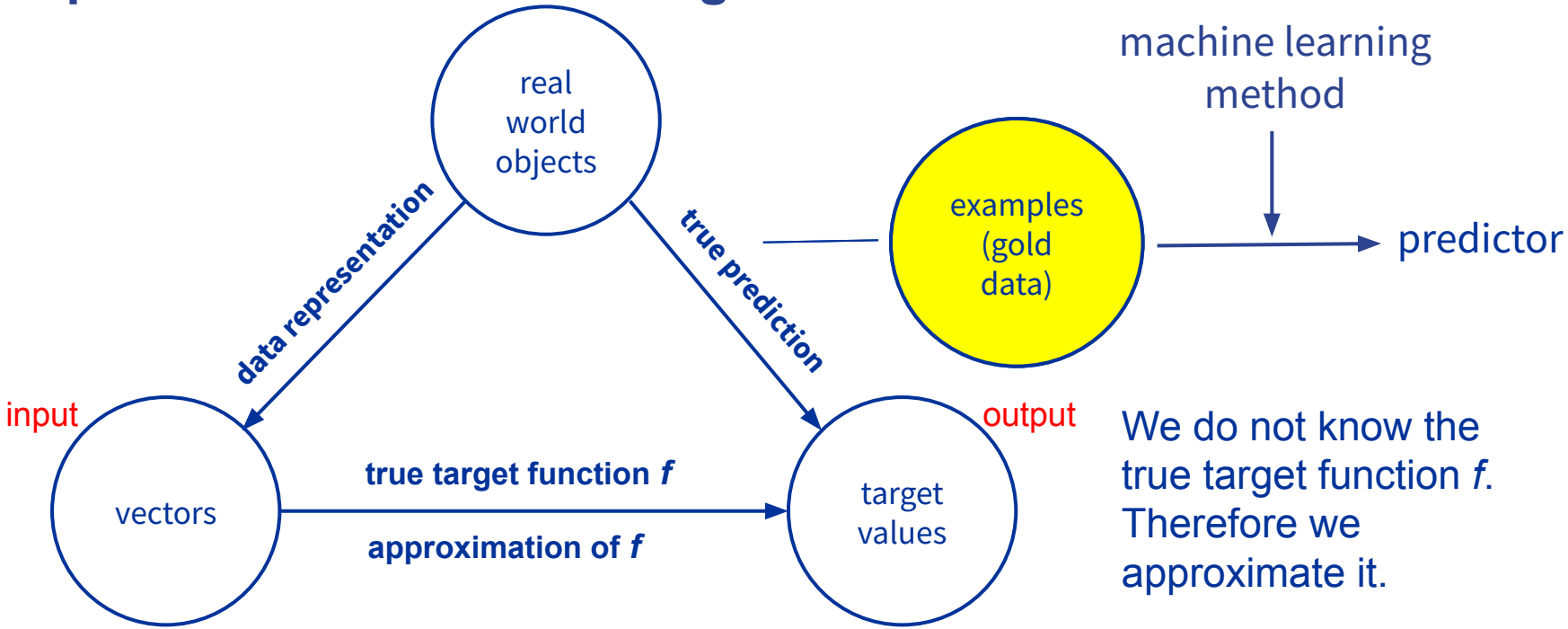
regression



Examples of Supervised machine learning

- Spam email detection,
  i.e. identify which new incoming emails are spam and which are not
- Regression analysis,
  e.g. a vehicle fuel consumption prediction based on its weight, and other attributes
- Image classification,
  e.g. classify images of animals as either cats or dogs
- Sentiment analysis,
  i.e. determine where a text is positive, negative or neutral
- Authorship detection,
  i.e. identify the author of a given text. It is a form of text classification that is often used in forensic linguistics, literary studies, and plagiarism detection

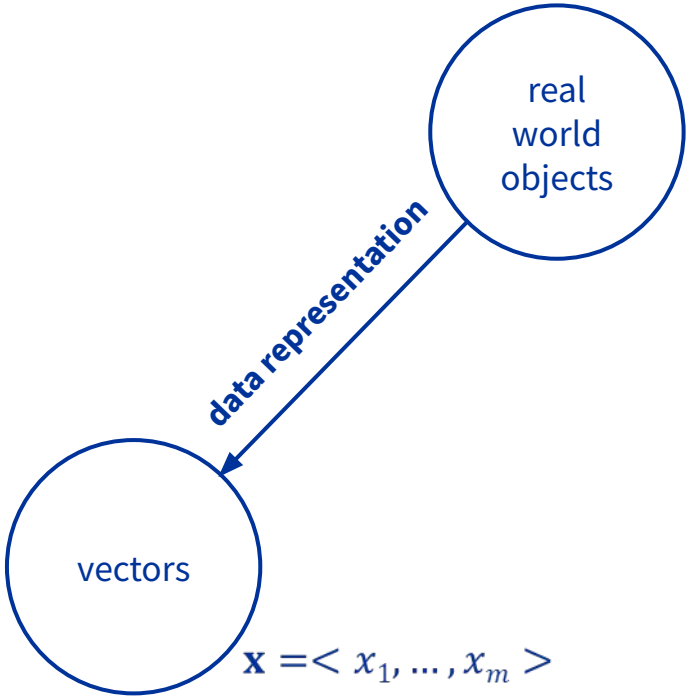
Supervised machine learning



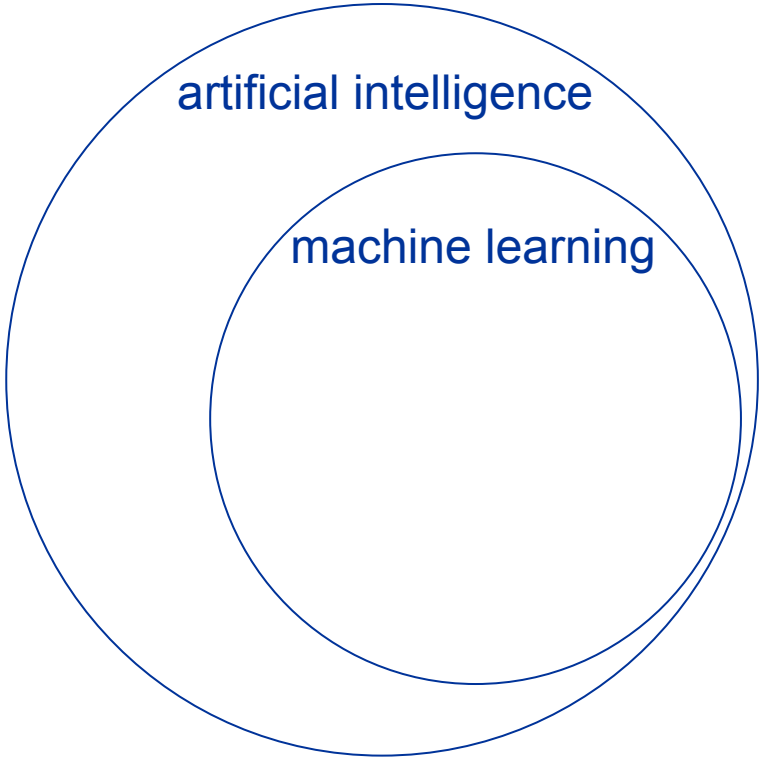
We do not know the true target function *f*. Therefore we approximate it.

Data representation

- readability by the computer
  - 1 example = 1 vector

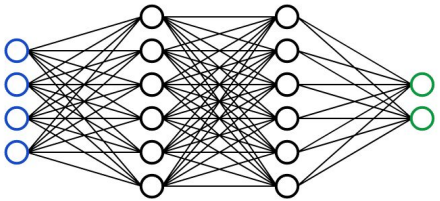


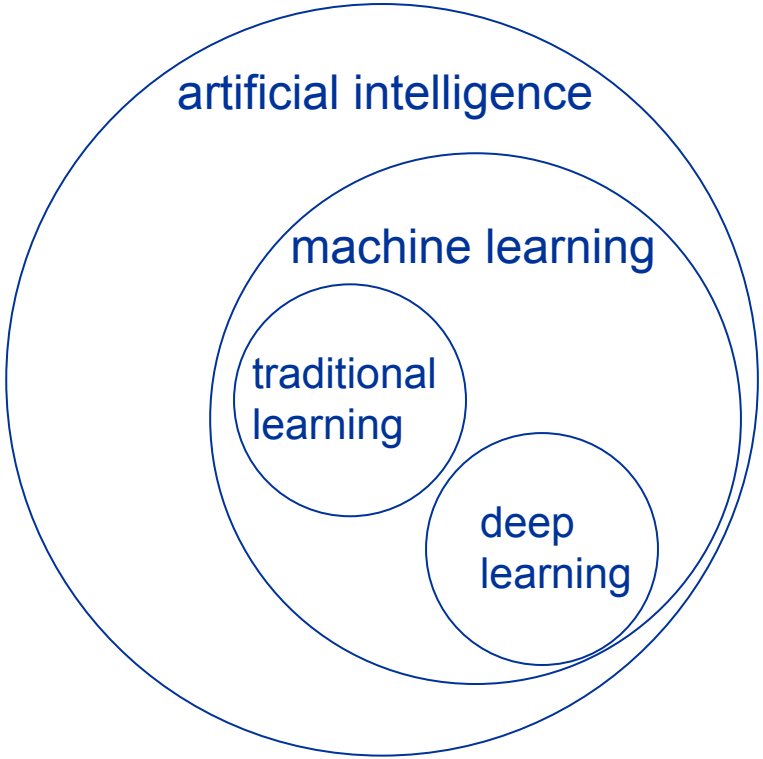
ML and Artificial Intelligence



ML and Deep learning

- A neural network consists of multiple layers of interconnected nodes, or neurons, that process input data and produce output data.
- Deep learning uses deep neural networks to model relationships in data.

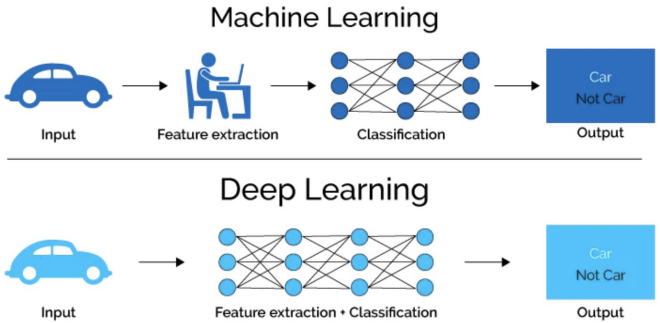




Traditional ML and Deep learning

Data representation

real world objects > vectors

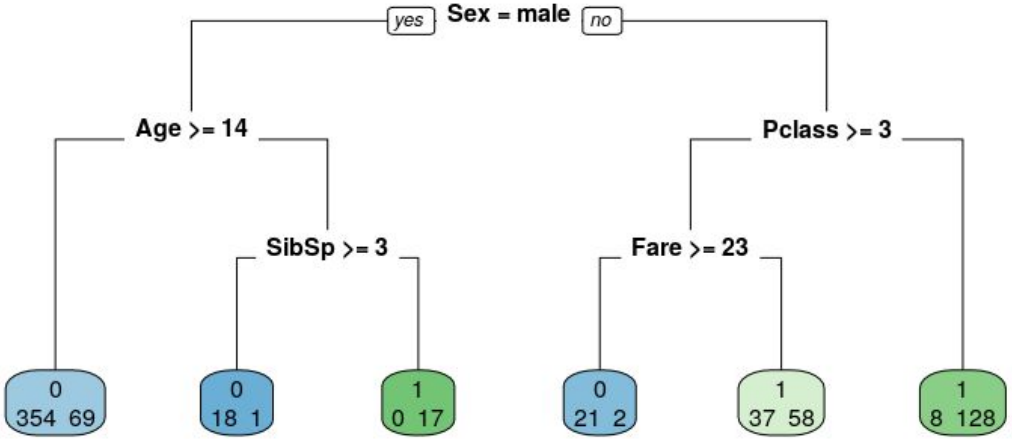


Recall Titanic dataset

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22.00	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.00	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.00	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.00	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35.00	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q

Decision tree learning algorithm

Model has a form of decision tree



Tree structure description

- Nodes
  - Root node
  - Internal nodes
  - Leaf nodes with target class values
- Decisions
  - Binary questions on a single feature, i.e. each internal node has two child nodes

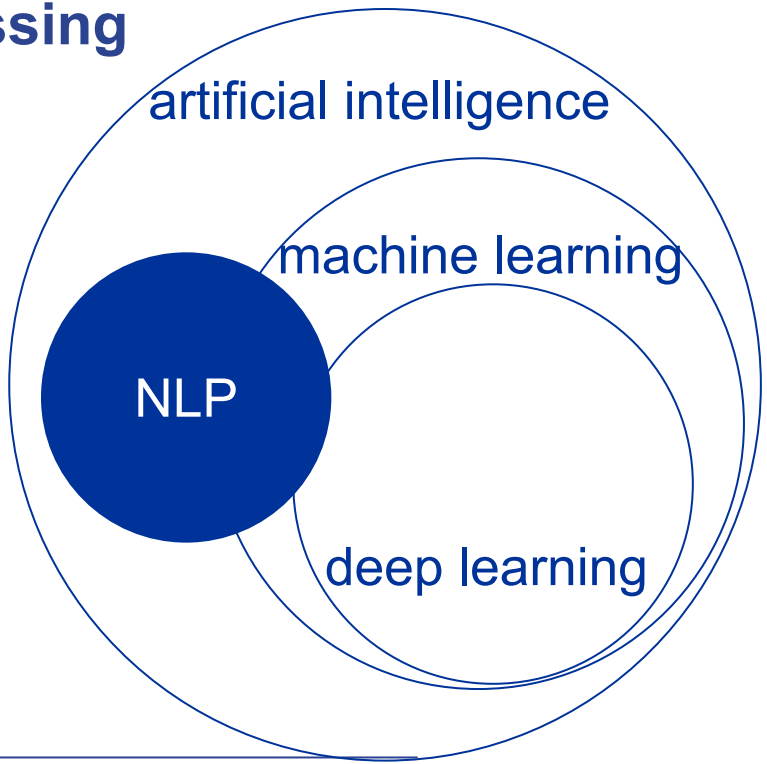
ML and Natural Language Processing

You've already met

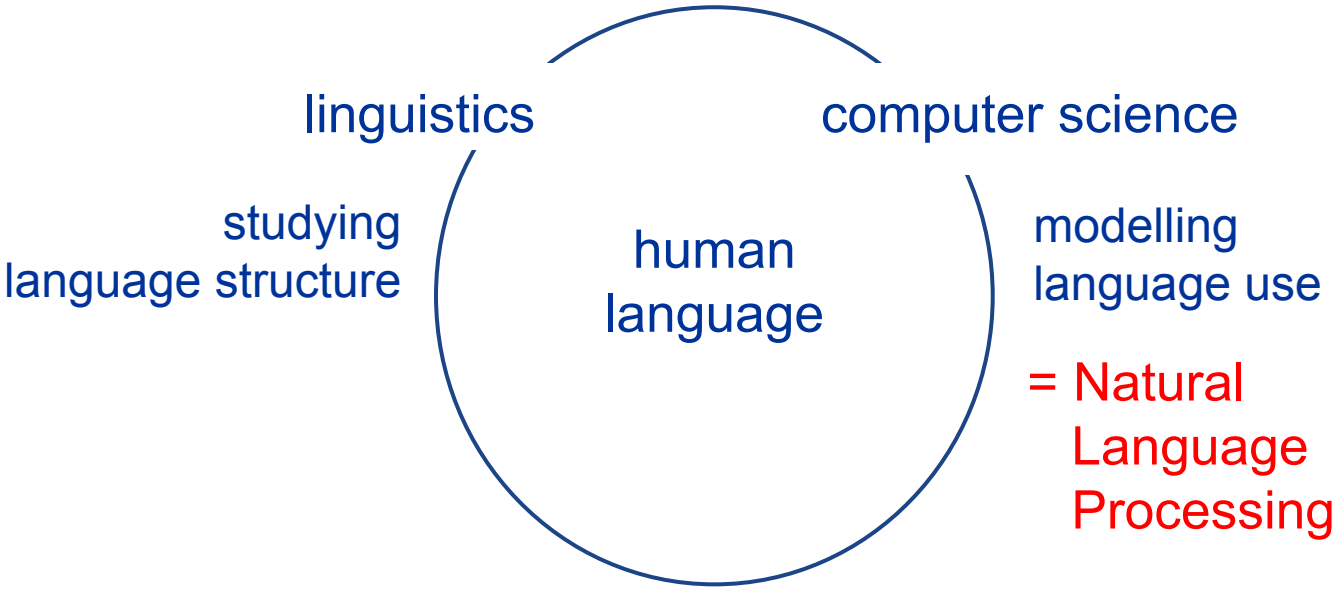
- UDPipe
- NameTag

see Lecture #6

These are the machine learning systems of Natural language processing (NLP).



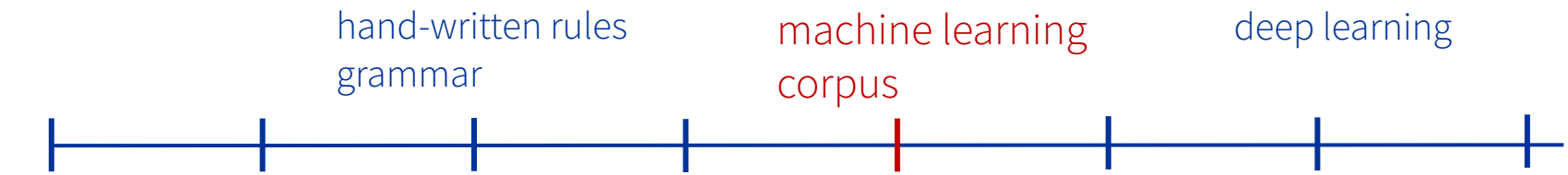
Computational linguistics



Natural Language Processing

- deals with computer and human interaction in both written and spoken natural language
- uses corpora of texts or speeches, sometimes enriched with linguistic information (annotated corpora)

see Lecture #6



rule-based machine translation (Weaver, 1949)

$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|F)$$

When we had our first results we submitted them to Coling 1988. Here is a part of the rejection review we received:

> The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30ff and references therein). The crude force of computers is not science. The paper is simply beyond the scope of COLING.
>
> Anonymous Coling review, 1 March 1988

Frederick Jelinek, ACL Lifetime Achievement Award, 2009

Evaluation of binary classifiers using a Confusion matrix

e.g. classification of Titanic survivors

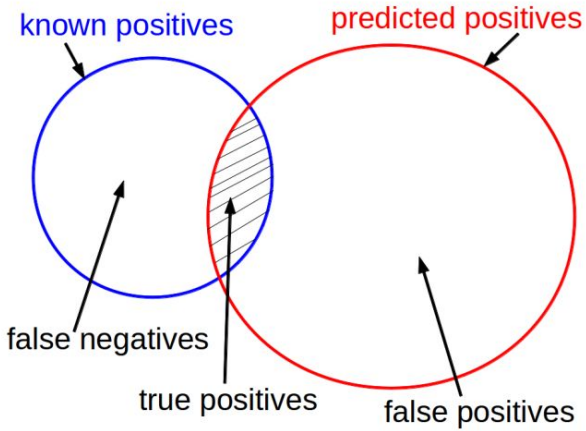
		Predicted class	
		Positive	**Negative**
True class	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Explanation

- **'Trues'** are examples correctly classified
- **'Falses'** are examples incorrectly classified
- **'Positives'** were predicted as positives (correctly or incorrectly)
- **'Negatives'** were predicted as negatives (correctly or incorrectly)

Confusion matrix :: Sensitivity TP/(TP+FN)

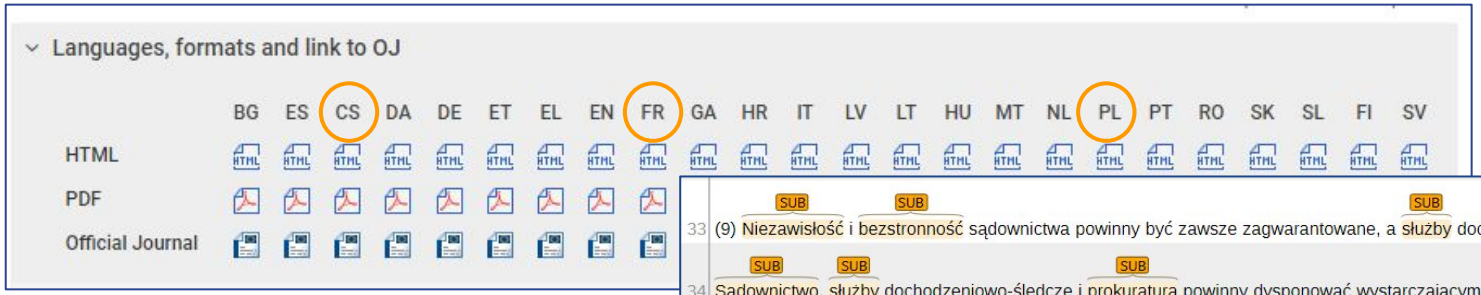
= ability of a classifier to correctly identify positive examples (e.g., survivors, patients with a disease)

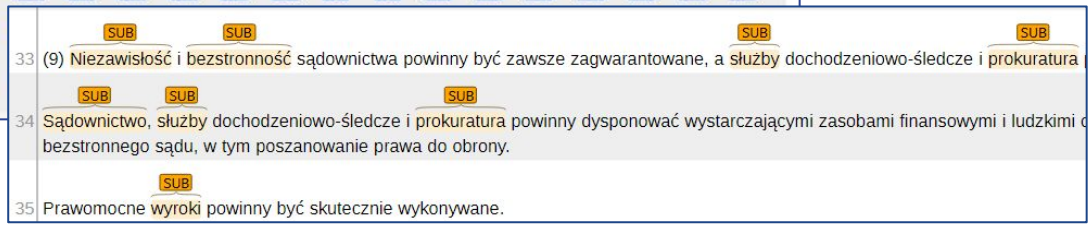


HW #6: Subject labeling of an EU regulation

- It is a classification task:
  identify which word in a text is or is not subject
- Regulation (EU, Euratom) 2020/2092 of the European Parliament and of the Council of 16 December 2020 on a general regime of conditionality for the protection of the Union budget, its preamble has 29 items
- https://ufal.mff.cuni.cz/courses/npfl134/preample

We already did this task with the CS, PL, FR versions, each version by 2 students, see Preamble 1.0





	CS	PL	FR
# of tokens	2,285	2,259	3,160

Inter-Annotator Agreement :: CS

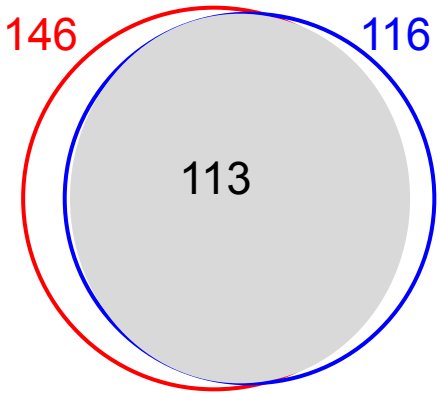
Confusion matrix for 2 students A1 and A2

A1 CS \ A2 CS	1	0	∑
1	113	33	146
0	3	2,136	2,139
∑	116	2,169	2,285

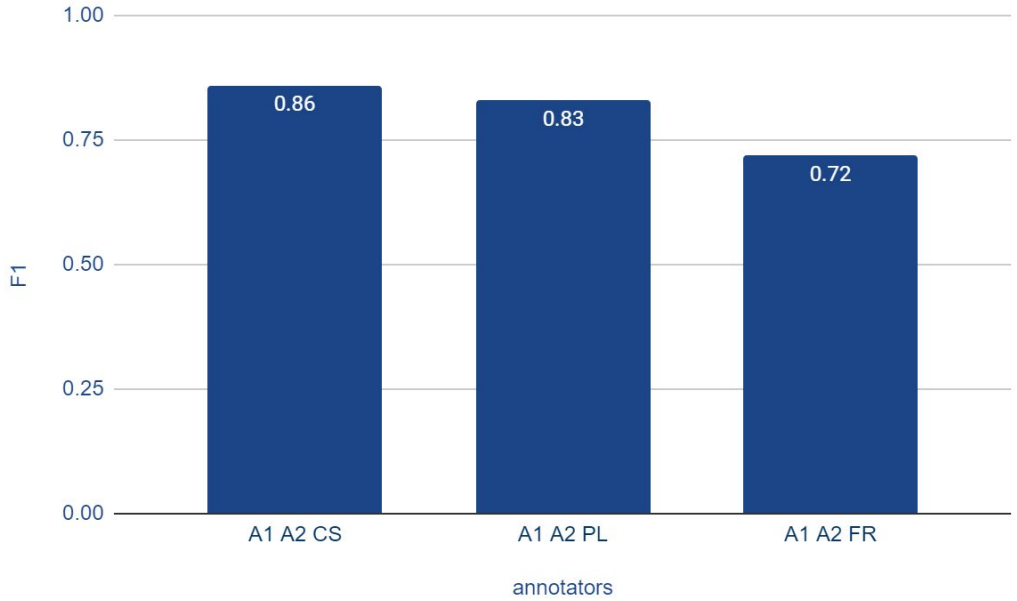
F1 = 2*0.97*0.77/(0.97+0.77) = 0.85

Precision = 113/(113+3) = 0.97

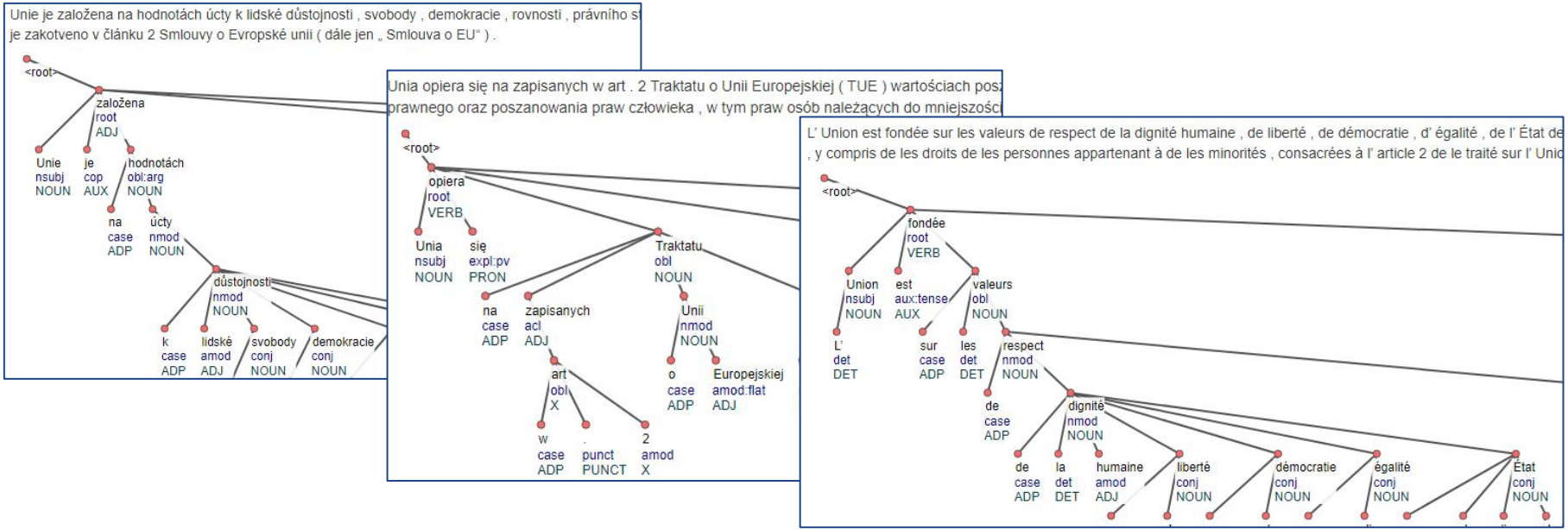
Recall = 113/(111+33) = 0.77



Inter-Annotator Agreement



What we also have - a complex syntactic analysis of each version using UDPipe



Evaluation of UDPipe vs. students