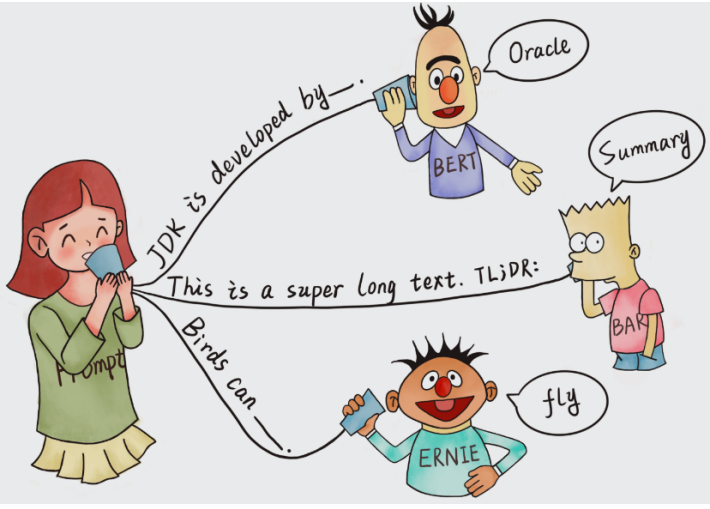
Basics

Introduction and Course Outline



Learning goals

- Understand the scope of the course
- Answers to all open question
- Get an impression of the workload

STRUCTURE OF THE COURSE

Central building blocks of the lecture

- Basic concepts (2 weeks)
- Transformer in-depth (1 week)
- BERT and subsequent models (2 weeks)
- T5 (1 week)
- GPT series and Prompting (2 weeks)
- More LLMs (1 week)
- Multilinguality (1 week)
- Math behind training LLMs (1 week)
- Current topics in research (1 week)

STRUCTURE OF THE COURSE

Central building blocks of the exercise

- Python essentials (week 1)
- Training an LSTM (week 2 – 3)
- Transformer (week 4 – 5)
- Fine-tuning T5 (week 6 – 7)
- GPT (week 8 – 9)
- ChatGPT / Multilinguality (week 10 – 11)

PREREQUISITES

Machine learning basics:

- A proper understanding of
  - linear algebra
  - loss functions
  - regularization
  - classification vs. regression
  - backpropagation and gradient descent
  - simple neural networks (MLPs)
- Great resource: I2ML course

MANAGING EXPECTIONS

What to expect

- A solid and proper unterstanding of
  - the central concepts and models of contemporary NLP
  - the capabilities and limitations of different models
- Challenging exercises that deepen your unterstanding, i.e.
  - you will have to invest quite some time in trying to solve them
  - just checking our solutions won't get you far
- An (inter)actively taught course with motivated lecturers, i.e.
  - we expect you to actively participate in the lecture and ask questions
  - you won't get much out of this course by just enrolling and not showing up

MANAGING EXPECTIONS

What not to expect

- A prompt engineering course
- Low workload and easy exercises (you will only learn this thoroughly be investing some time)
- Only the latest state-of-the-art LLMs (you also need to know the basics)

NOTATION (IMPORTANT!!)

$\mathcal{C}$	A corpus of M documents
$\mathbf{C}$	The confusion matrix of a classification problem
C	Size of the context window
E	Embedding size // Size of the vector representation
h	Number of Attention heads in a Transformer-based architecture
H	Dimension of the hidden layer of a neural network
i	Running index for observations in a data set, words/tokens in a sequence or documents in a corpus
j	Running index for covariates or levels of a categorical target variable
k	The number of classes of a categorical target variable
$\mathcal{L}$	A data generating process a.k.a. language
$\mathcal{M}$	A (language) model

NOTATION (IMPORTANT!!)

m	The number of words/tokens in a sequence s
M	The number of documents in a corpus $\mathcal{C}$
n	The number of observations in a data set
p	The number of features in a tabular data set
s	An ordered sequence of m words
θ	The parameters of a model
$V = \{w_1, \ldots, w_N\}$	The vocabulary of a corpus $\mathcal{C}$ or a language $\mathcal{L}$
$\lvert V \rvert$ or N	The size of a vocabulary V
$w_{[t]}$	The word/token at the t-th position in a sequence
$\vec{w}^{(\text{word})}$	vector representation of a word
$\mathbf{W}$	The weight matrix of a neural network
x	Influential variable(s)
y	Target variable(s)

Any Questions?