Introduction to language modelling

Marco Kuhlmann
Department of Computer and Information Science

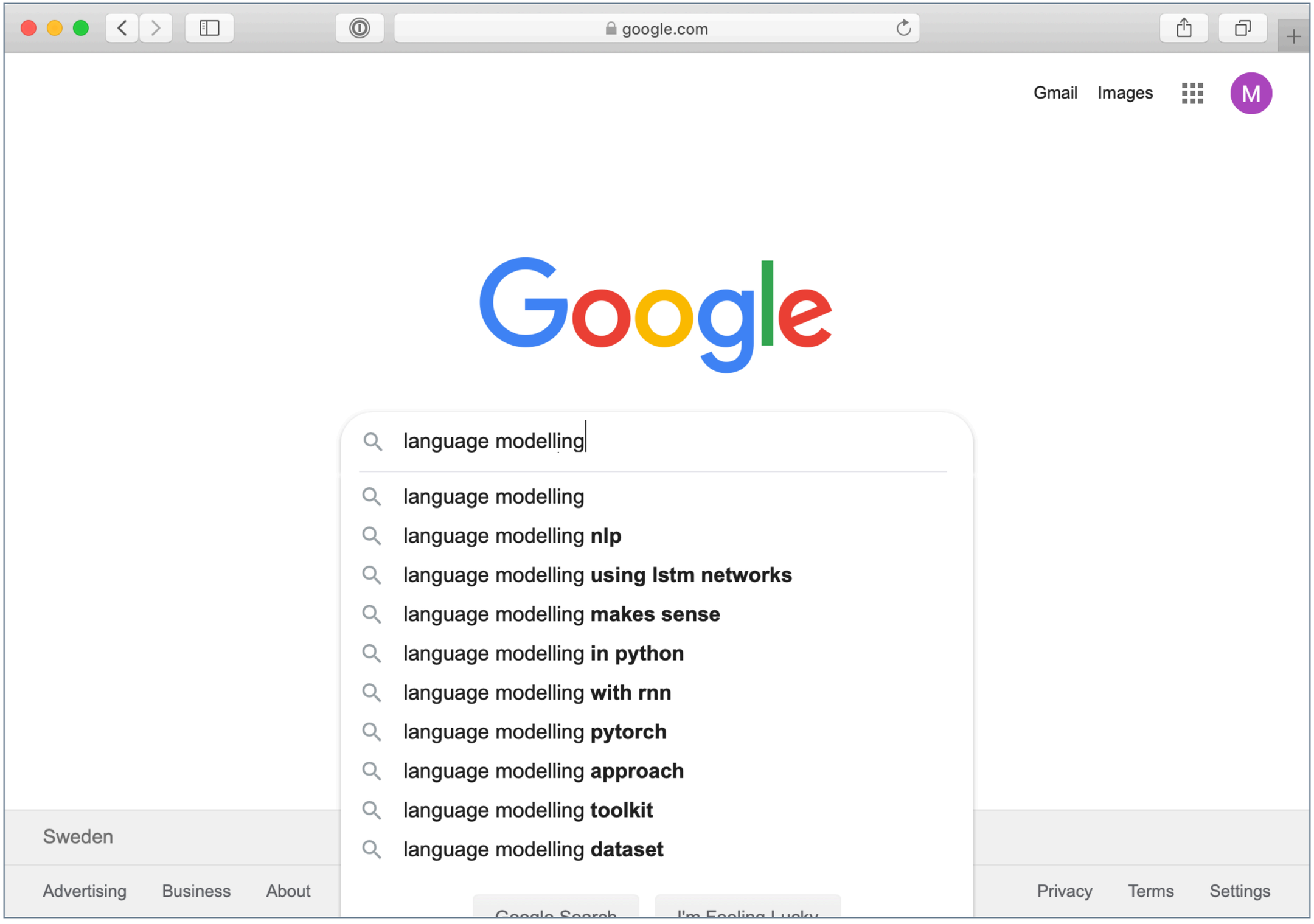
Language modelling

- **Language modelling** is the task of predicting which word comes next in a sequence of words.
- More formally, given a sequence of words $w_1, \ldots, w_t$, we want to know the probability of the next word, w_{t+1}:

$$P(w_{t+1} \mid w_1, \ldots, w_t)$$

- Here we are assuming that w_{t+1} comes from a fixed vocabulary V.

  This allows language modelling to be treated as a classification task.


google.com
Gmail
Images
M
Google
language modelling
language modelling
language modelling nlp
language modelling using lstm networks
language modelling makes sense
language modelling in python
language modelling with rnn
language modelling pytorch
language modelling approach
language modelling toolkit
language modelling dataset
Sweden
Advertising
Business
About
Privacy
Terms
Settings

An alternative view on language models

- Rather than as predictive models, language models can also be viewed as models that assign probability to a piece of text.

  How likely is it that this piece of text is written in Swedish? French?

- These two views are equivalent, as the probability of a sequence can be expressed as a product of conditional probabilities: *

$$P(w_1 \cdots w_N) = \prod_{t=1}^{N} P(w_t \mid w_1, \ldots, w_{t-1})$$

* There are some subtleties here that we gloss over.

Language models for translation

Rank candidate translations for the following Chinese sentence:

他 向 记者 介绍了 主要 内容

he to reporters introduced main content

- he introduced reporters to the main contents
- he briefed to reporters the main contents
- he briefed reporters on the main contents

Example from Jurafsky and Martin (2017)

talktotransformer.com

Private Research Teaching LiU

Talk to Transformer

See how a modern neural network completes your text. Type a custom snippet or try one of the examples. Learn more below.

Follow @AdamDanielKing for more neat neural networks.

Custom prompt

Kim hates broccoli.

GENERATE ANOTHER

Completion

Kim hates broccoli. They're making her sick. Well, it looks like the other week she had a flashback and went into the freezer where all of the vegetables used to be and she has never seen a broccoli plant before, so she never knew that if you kept it out of the cold it would magically grow to enormous size. Eventually she saw that it grew. So basically, thanks to a self-imposed dorm shortage in the lower four-fours, I have recently arranged to fill the highest-

Evaluating language models

Eisenstein § 6.4

- **Intrinsic evaluation**

  How does the method or model score with respect to a given evaluation measure?

  examples from text classification: precision and recall

- **Extrinsic evaluation**

  How much does the method or model help the application in which it is embedded?

  predictive input, machine translation, speech recognition

Perplexity

Eisenstein § 6.4

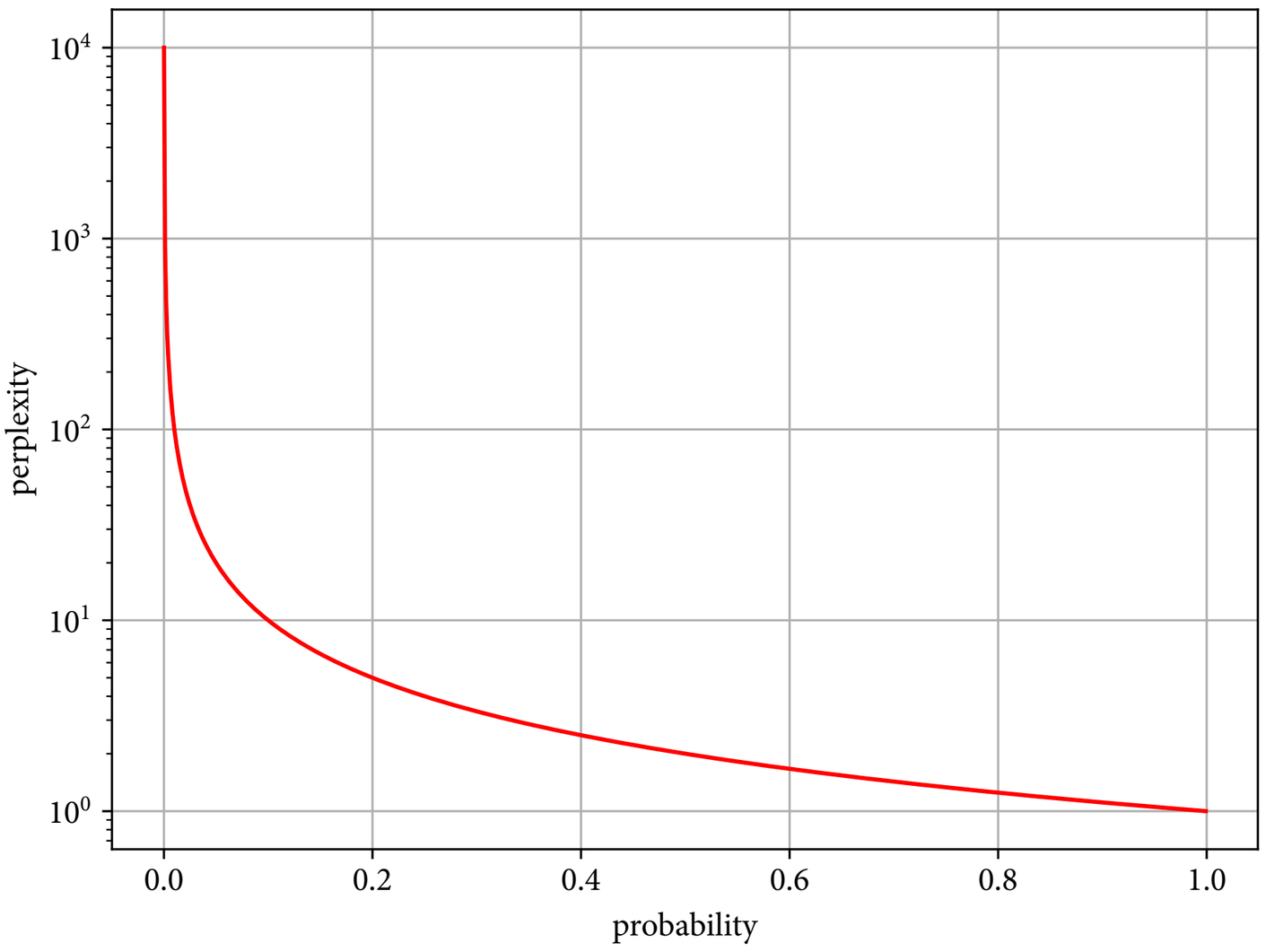
- Intrinsic evaluation of language models is based on the likelihood that a model assigns to held-out data.

- Formally, we compute the cross-entropy between two probability distributions: a language model and the empirical distribution.

- This cross-entropy is usually presented as **perplexity**:

$$2^{-\frac{1}{N}\log_2 P(w_1 \cdots w_N)}$$

Perplexity



Perplexity benchmarks

Model	Perplexity
Random guessing	33,278
Fixed-window bigram model (lab), 1 epoch	344
Fixed-window bigram model (lab), 2 epochs	333
Interpolated trigram model (advanced lab)	298
Recurrent neural network model (lab), 4 epochs	242
State-of-the art model (35M parameters)	40

Results on the validation section of the WikiText-2 dataset. Lower is better.