Sequence Learning

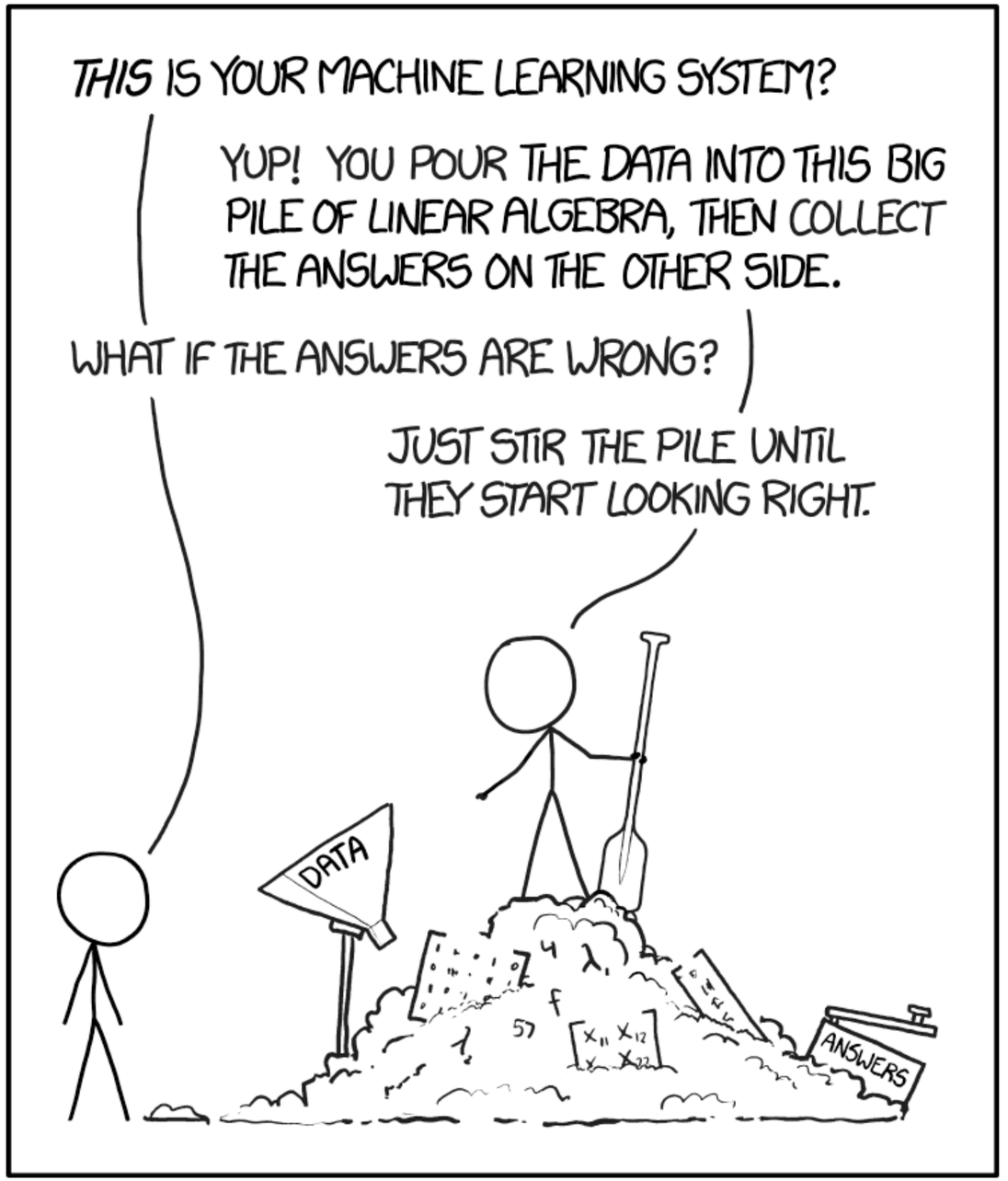
Introduction

Korbinian Riedhammer

TECHNISCHE HOCHSCHULE NÜRNBERG
GEORG SIMON OHM

Today

- Logistics
- Why I teach this class
- Why you should take this class
- Motivation
- Syllabus
- What you should bring to this class



https://xkcd.com/1838/

Logistics

- Mondays at 9.45a SP.467, discussion on Teams (Code: y6n8dbx)
- Materials: https://seqlrn.github.io (continuously updated…)
- Exam:
  - mandatory assignments in python (pair-programming ok; individual submissions required)
  - 20’ oral exam in the last week of lecture period (calendar week 27)

Why I teach this class

- Industry background in speech recognition/indexing (mod9.io)
- Research focus
  - Speech processing for medical applications (eg. stuttering, dementia)
  - Speech recognition for indexing/search
  - Sequence learning for industrial applications (mostly anomaly detection)

Why you should take this class

- Machine learning is the future*
- Many applications are to sequences, not single observations
- Understand the foundations of sequence classification

*or at least a very well paid part of it

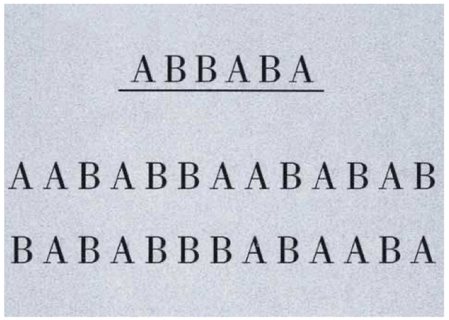
Flashback: Verbmobil

Research project 1993-2000 (!)

https://www.youtube.com/watch?v=DcG9-KWx0Fg

Memory Clinic

Cooperation with Klinikum Nürnberg



- Can we use speech processing to automate dementia tests?

Table 1: *Automated SKT scoring on manual transcriptions (Trans.) and automatic speech recognition with (ASR-5) and without (ASR-1) the top five word alternatives. Column Top-21 refers to top 21 speakers and ASR-5.*

ID	Test/Task	Trans.	ASR-1	ASR-5	Top-21
1	naming objects	0.89	0.70	0.81	0.89
2	reproducing objects	1.00	0.58	0.71	0.83
3	reading numbers	0.94	0.85	0.86	0.94
6	counting symbols	0.90	0.59	0.58	0.54
7	interference test	0.99	0.97	0.98	0.99
8	naming after distraction	1.00	0.75	0.90	0.97
9	recognizing objects	0.89	0.50	0.55	0.68
-	attention score	0.92	0.84	0.82	0.85
-	memory score	0.98	0.62	0.78	0.93
-	total score	0.97	0.81	0.89	0.94

Table 2: *Automated CERAD scoring on manual transcriptions (Trans.) and automatic speech recognition with (ASR-5) and without (ASR-1) the top five word alternatives. Column Top-21 refers to top 21 speakers and ASR-5.*

ID	Test	Trans.	ASR-1	ASR-5	Top-21
1	verbal fluency test	0.98	0.82	0.85	0.91
2	Boston Naming Test	0.70	0.14	0.24	0.47
3	MMSE	0.71	0.07	0.35	0.52
4	word list learning	0.94	0.62	0.70	0.75
6	word list recall	0.99	0.68	0.81	0.78
-	total	0.71	0.37	0.49	0.61

Data Sources

Analog signals (discretized)

- Microphones
- Vibrations
- Conductivity
- Ambient: pressure, temperature, humidity, …
- Positional: GPS, gyro, distances
- User input: key-press, gestures, pressure, swipe, …

Data Sources

Digital or “Big Data” signals

- Text 🫣
- Log streams
- Network traffic
- Events (IoT, MQTT, …)
- User-generated content (Twitter, blogs, …)

Tooling

- python3
- jupyter
- numpy/scipy
- PyTorch
- 🤗 Transformers

Syllabus

- Basic algorithms
  - Matching and comparing (discrete) sequences, Dynamic programming
- Statistical modeling
  - Markov chains, hidden Markov models
  - Maximum likelihood, expectation maximisation
- Neural networks
  - Feed-forward and recurrent networks
  - Attention and transformers
  - Transfer learning
- Reinforcement learning

Assignments

- Jupyter notebooks for every chapter
- Submission mandatory (but not graded)
- Programming
- Evaluation
- Transfer to similar tasks or data sets

What you should bring to this class

- A little bit of probability theory
- A little bit of optimization theory
- Algorithms and programming
- Curiosity and perseverance: understanding is hard, implementing sometimes even harder…