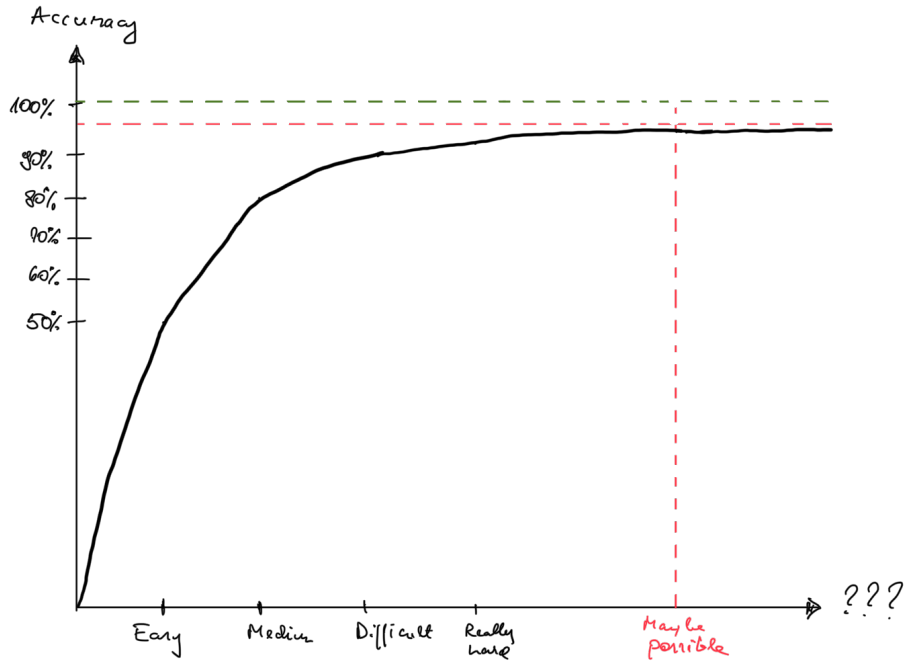
LTAT.02.004 Machine Learning II

Introduction

Sven Laur
University of Tartu

Why there are no self-driving cars?



- ▷ Gathered data, Background knowledge
- ▷ Computational resources, Learning algorithms
- ▷ Uncanny valley, Public perception

Data gathering is programming

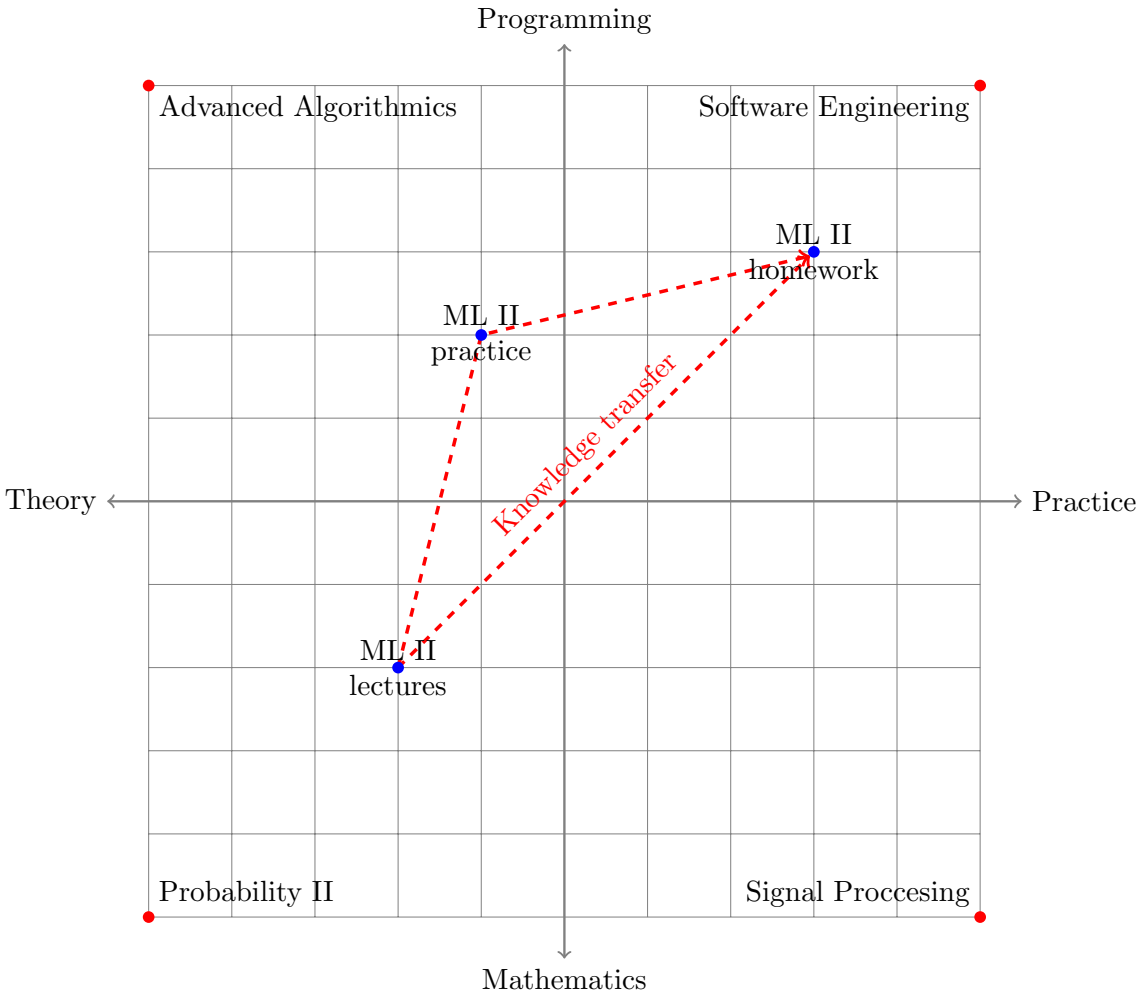
How to increase volume

- ▷ Self-labelled data vs manual labelling
- ▷ Symmetries and data augmentation
- ▷ Consistency and data augmentation
- ▷ Backgroung knowledge and data augmentation

Software engineering practices

- ▷ Version control $\rightsquigarrow$ Data and model versioning
- ▷ Code differences (blame) $\rightsquigarrow$ Model comparison
- ▷ Bugtracking $\rightsquigarrow$ Model diagnostics & Abalation studies
- ▷ Tests $\rightsquigarrow$ Dedicated testsets
- ▷ Automatic testing $\rightsquigarrow$ Robustness analysis and Advdersarial learning

What is this course about?



Course plan

- ▷ **Performance evaluation**
  ⇝ absolute risk, relative risk, CLT, crossvalidation, bootstrap
- ▷ **Probabilistic modelling**
  ⇝ frequentism, bayesianism, statistical tests, confidence intervals
- ▷ **Sequence models**
  ⇝ Markov chain, Hidden Markov Mode, belief propagation
- ▷ **Multivariate normal distribution**
  ⇝ Models behind linear regression and linear time series analysis
- ▷ **Affine transformations**
  ⇝ Blind Source Separation, PCA, LDA, ICA, NMF
- ▷ **Model based clustering**
  ⇝ Hierarchical clustering, k-means, gaussian mixture model
- ▷ **Expectation-maximisation algorithm**
  ⇝ weighted data, gaussian mixture model
- ▷ **Expectation-maximisation algorithm for sequential models**
  ⇝ Hidden Markov Model, Kalman filter