

# IB031: Úvod do strojového učení

---

Tomáš Brázdil, Luboš Popelínský, Jaroslav Čechák &  
his boys

Kdo s kým, o čem, proč

Kdo s kým, o čem, proč

# Kdo s kým, o čem, proč

## Outline

- ▶ Supervised learning
- ▶ Learning decision trees
- ▶ Evaluation
- ▶ Pre-processing
- ▶ Clustering; Lazy learning
- ▶ Anomaly detection
- ▶ On machine learning theory
- ▶ Probabilistic classifiers
- ▶ Linear models
- ▶ Kernel Methods
- ▶ Neural nets

**Teaching materials:** ISMU

Kdo s kým, o čem, **proč**

# Organizace

- ▶ přednášky
- ▶ cvičení 2h
- ▶ projekt
- ▶ semestrální zkouška
- ▶ písemná zkouška

## Závěrečné hodnocení

- ▶ semestrální zkouška 25b.
- ▶ projekt 30b. (min 15b.)
- ▶ závěrečná zkouška 45b. (min 15b.)
- ▶ <50 F, <60 E, <70 D, <80 C, <90 B, >=90 A  
40 zápočet

# Co je strojové učení

Herbert Simon (1960s): “Learning is any process by which a system improves performance from experience.”

Tom Mitchell (1990s, paraphrased): “Learning aims to improve task, T, with respect to performance metric, P, based on experience, E.”

# Co je strojové učení

Herbert Simon (1960s): “Learning is any process by which a system improves performance from experience.”

Tom Mitchell (1990s, paraphrased): “Learning aims to improve **task, T**, with respect to **performance metric, P**, based on **experience, E**. ”



# Příklady

T: Playing checkers

P: Percentage of games won against an arbitrary opponent

E: Playing practice games against itself

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.

P: Percentage of email messages correctly classified.

E: Database of emails, some with human-given labels

# Příklady

T: Playing checkers

P: Percentage of games won against an arbitrary opponent

E: Playing practice games against itself

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.

P: Percentage of email messages correctly classified.

E: Database of emails, some with human-given labels

# Příklady

T: Playing checkers

P: Percentage of games won against an arbitrary opponent

E: Playing practice games against itself

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.

P: Percentage of email messages correctly classified.

E: Database of emails, some with human-given labels

# Příklady

T: Playing checkers

P: Percentage of games won against an arbitrary opponent

E: Playing practice games against itself

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.

P: Percentage of email messages correctly classified.

E: Database of emails, some with human-given labels

Další příklady?

# Třídy úloh

- ▶ shlukování
- ▶ klasifikace a predikce
- ▶ hledání asociací
- ▶ detekce anomálií

# Historie

- ▶ 1950s :
  - Alan Turing and NP-hard problems
  - Samuel's checker player, see Ray Mooney ML Course slides
- ▶ 1960s :
  - Neural networks: Perceptron
  - Pattern recognition
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron
- ▶ 1970s :
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck;
  - Scientific discovery with BACON and AM (math)
  - Quinlan's ID3
  - Michalski's AQ

# Historie

- ▶ 1980s :
  - Advanced decision tree and rule learning
  - Learning and planning and problem solving
  - Resurgence of neural networks (connectionism, backpropagation)
  - Valiant's PAC Learning Theory
  - Focus on experimental methodology
- ▶ 1990s :
  - Data mining
  - Text learning
  - Reinforcement learning (RL)
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning
  - Web mining
  - Weka



# Historie

- ▶ 2000s :
  - Support vector machines. Kernel methods
  - Statistical relational learning
  - Graph and Sequence mining, Link learning
  - Privacy-preserving data mining
  - Security (intrusion, virus, and worm detection)
  - Recommender systems; Personalized assistants that learn
  - Visual data mining
  - Stream mining
  - RapidMiner
  - R for machine learning

# Historie

- ▶ 2006 :  
Deep learning
- ▶ 2010s :  
KNIME  
Big data, Big data, Big data . .  
Outlier detection and explanation  
Automated machine learning  
Deep learning in practice