Decision-theoretic Machine Learning

Krzysztof Dembczyński and Wojciech Kotłowski

Intelligent Decision Support Systems Laboratory (IDSS) Poznań University of Technology, Poland



Poznań University of Technology, Summer 2015

• Wojciech Kotłowski:

- Main interest: machine learning (online-learning, matrix learning, preference learning, learning theory)
- Past interest: multi-criteria decision aiding, decision rule models
- ► Education: MSc (2004 in Computer Science, 2006 in Physics), PhD (2009).
- ► Postdoc: Centrum Wiskunde & Informatica (2009-2012).



• Krzysztof Dembczyński:

- Main interest: machine learning (multi-label classification, structured output prediction, preference learning, learning theory)
- Past interest: multi-criteria decision aiding, decision rule models
- Education: MSc (2001), PhD (2009).
- ▶ Postdoc: Marburg University (2009-2011).



- We both work in Laboratory of Intelligent Decision Support Systems, headed by Roman Słowiński, at Poznań University of Technology.
- We try to do our best in machine learning: conference articles at ICML, NIPS, COLT, journal articles in JMLR, MLJ, DMKD, TDE, serving as PC members of major AI conferences (ICML, NIPS, COLT, IJCAI, AAAI, KDD).

We live in the era of Big Data and Machine Learning.



Search engines: website ranking and personalization



Recommender systems: movie, book, product recommendations



Autonomous vehicles



Spam filtering

kaggle

Customer Solutions Competitions

ons Community -

Welcome to Kaggle, the leading platform for predictive modeling competitions. Here's how to jump into competing on Kaggle —

New to Data Science? Visit our Wiki » Learn about hosting a competition » in-Class & Research competitions »

₽∎±

Enter

Find a competition & download the training data. You don't need new software/skills to submit. 9 ¢ ılı Build

Build a model using whatever methods you prefer and upload your predictions to Kaggle.

Kaggle scores your solution in real time and you'll see your place on the live leaderboard.

Active Competitions	Active Competitions			
All Competitions	Ŧ	_	Acquire Valued Shoppers Challenge Predict which shoppers will become repeat buyers	16 days 901 teams \$30,000
		AVITO	The Hunt for Prohibited Content Predict which ads contain illicit content	3 months 28 teams \$25,000

A plenty of machine learning competitions

• What is machine learning?

- What is machine learning?
 - Machine learning is the science of getting computers to act without being explicitly programmed. (Andrew Ng)

- What is machine learning?
 - ► Machine learning is the science of getting computers to act without being explicitly programmed. (Andrew Ng)
- Supervised learning
 - ► Learn a computer to predict an unknown response/value of a decision attribute for a given object described by several features.

- What is machine learning?
 - ► Machine learning is the science of getting computers to act without being explicitly programmed. (Andrew Ng)
- Supervised learning
 - Learn a computer to predict an unknown response/value of a decision attribute for a given object described by several features.
- Two main supervised problems:
 - ► Classification: Prediction of categorical response,
 - Regression: Prediction of continuous response.

- What is machine learning?
 - ► Machine learning is the science of getting computers to act without being explicitly programmed. (Andrew Ng)
- Supervised learning
 - Learn a computer to predict an unknown response/value of a decision attribute for a given object described by several features.
- Two main supervised problems:
 - ► Classification: Prediction of categorical response,
 - ► Regression: Prediction of continuous response.
- Examples:
 - ► Spam filtering,
 - Handwriting recognition,
 - Text classification,
 - Stock prices,
 - ► etc.

- We know relatively much about solving simple learning problems such as binary classification:
 - Advanced theory,
 - Implemented fast algorithms,
 - Almost a mature technology.
- The main challenges are:
 - ► feature engineering,
 - supervision of examples,
 - new applications,
 - new complex problems,
 - large-scale problems.

The aim and scope of this lecture

• Aim: To explain theoretical foundations of machine learning in order to show how simple algorithms can be used for solving complex problems.

• Scope:

- 1 Introduction to Machine Learning
- 2 Binary Classification
- 3 Bipartite Ranking
- 4 Multi-Label Classification

Bibliography

- T. Hastie, R. Tibshirani, and J.H. Friedman. *Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, second edition, 2009 http://www-stat.stanford.edu/~tibs/ElemStatLearn/
- Christopher M. Bishop. Christopher M. Bishop. Springer-Verlag, 2006
- David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012 http://www.cs.ucl.ac.uk/staff/d.barber/brml/
- Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning From Data*. AMLBook, 2012

Software

- Weka (http://www.cs.waikato.ac.nz/ml/weka/)
- R-project (http://www.r-project.org/),
- Octave (https://www.gnu.org/software/octave/),
- Julia (http://julialang.org/),
- Scikit-learn (http://scikit-learn.org/stable/)
- Matlab (http://www.mathworks.com/products/matlab/)
- H20 (http://0xdata.com/)
- GraphLab (http://dato.com/)
- MLLib (https://spark.apache.org/mllib/)
- Mahout (http://mahout.apache.org/)

• . . .