Learning for Learning
Machine Learning to Support Education

Christian Igel
Department of Computer Science
Outline

1. Machine Learning & Data Mining
2. School Drop Out Prediction
3. Machine Learning at DIKU
Outline

1 Machine Learning & Data Mining

2 School Drop Out Prediction

3 Machine Learning at DIKU
Machine learning

Machine learning research studies software that improves its performance at a given task based on sample data or experience.
Supervised machine learning

Learning is goal-directed changing of behaviour based on experience.

Supervised machine learning automates the process of *inductive inference*:

1. Observe a phenomenon
2. Construct a model of that phenomenon
3. Make predictions using this model
Example: Sport analytics

Learning by heart is not enough

- Good performance on data used for building a model does not imply good generalisation (i.e., good performance on patterns not seen during training).

- Overfitting: The hypothesis faithfully reflects idiosyncrasies of the training data rather than the underlying distribution.
Outline

1. Machine Learning & Data Mining
2. School Drop Out Prediction
3. Machine Learning at DIKU
School drop out

- Next slides based on DIKU MSc thesis by Nicolæ-Bogdan Şara in collaboration with MaCom
- School drop out is a problem for the individual and society
- Drop out rates in Denmark in the year 2009:
  - STX education: 17% drop out
  - HF education: 25% drop out
- 24% drop out rate in MaCom database
- Idea: Apply machine learning algorithms for assessing whether high-school students will drop-out in the next 3 months
Data acquisition

1 year sample: grades, absences, assignments. Ignore vacations

Special case with students having missing or different data which may require different classifier and/or features

Year 1 end

Random 1 month sample: March, September of November. Use grades, absences, assignments and variation. Ignore vacations. In case of dropout student this is not random but taken exactly before the 3 months period.

Year 2 end

Year 3 end

Graduate student timeline

Drop out student timeline

3 months period where the student might drop out. No data is extracted from this.

Drops out here
Performance results

Classifiers in ROC space

- Random Forest
- CART
- Naïve Bayes
- SVM

True Positives Rate vs. False Positives Rate
Indicators of drop out

Count features at nodes used in model construction for RF
Outline

1. Machine Learning & Data Mining
2. School Drop Out Prediction
3. Machine Learning at DIKU
DIKU researchers in learning systems

Machine Learning Lab
http://image.diku.dk/MLLab
Image Section
http://www.diku.dk/english/research/imagesection

DIKU faculty doing machine learning, information retrieval, and pattern recognition: Corinna Cortes (head of Google Research New York, adjunct), Ingemar Cox, Marleen De Bruijne, Sune Darkner, Fabian Gieseke, Aasa Feragen, Christian Igel (head of ML Lab), Francois Lauze, Christina Lioma, Mads Nielsen (head of Image Group), Marco Loog (TU Delft, adjunct), Søren Olsen, Yevgeny Seldin, Jon Sporring, Kim Steenstrup Pedersen, ...
When theory and practice meet . . .

Making machine learning systems applicable by increasing their autonomy and scalability using concepts from computational and optimisation theory and from biological information processing.

http://image.diku.dk/MLLab
http://image.diku.dk/igel

Pellecchia et al.: *IEEE Intelligent Sys* 20, 2005
Suttrop, Igel: *Multi-objective Machine Learning* Ch. 9, Springer, 2006

Igel et al.: *IEEE/ACM TCBB* 4, 2007
Mersch et al.: *IJNS* 17, 2007

Mayr et al.: *Analytical Chemistry* 75, 2003