National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
Institute of Physics and Technology

Lecture 1
Introduction to Machine Learning

Dmytro Progonov,
PhD, Associate Professor,
Department Of Physics and Information Security Systems

Content

- Machine learning – what and why?
- Types of machine learning;
- Practical applications and open-problems in machine learning;
- System identification task.

Problem statement

- ***2.7 Zetabytes*** (2.7×10^{21}) of data exist in the digital universe today;
- ***235 Terabytes*** of data has been collected by the U.S. Library of Congress in April 2011;
- More than ***5 billion people*** are calling, texting, tweeting and browsing on mobile phones worldwide;
- YouTube users upload ***48 hours*** of new video ***every minute*** of the day;
- Every person in the world having more than ***215m high-resolution MRI scans*** a day;
- ***294 billions emails*** sent ever day;
- ***Trillions of sensors*** monitor, track and communicate with each other, populating the Internet-of-Things with real-time data.

Machine learning

Machine learning provides automated methods of data analysis. In particular, we define ***machine learning*** as a set of methods that can *automatically detect patterns* in data, and then use the uncovered patterns to *predict future data*, or to perform other kind of *decision making under uncertainty*.

Types of machine learning (1/2)

The ***predictive*** or ***supervised learning*** – the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y, given a labeled set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$;

The ***descriptive*** or ***unsupervised learning*** – the goal is to find "interesting patterns" in unlabeled data $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^{N}$;

Reinforcement learning – is used for learning how to act or behave when given occasional reward or punishment signals.

Types of machine learning (2/2)

Supervised learning

- Classification
- Regression

Unsupervised learning

- Knowledge discovery
- Density estimation
- Cluster discovering
- Latent factor discovering
- Matrix completion
- System identification

Reinforcement learning

- Learning and acting under partial information
- Transfer learning
- Large-scale empirical evaluations

Practical applications

Supervised learning

- Document classification and email spam filtering
- Classifying flowers
- Image classification and handwriting recognition
- Face detection and recognition
- Predict tomorrow's stock market state

Unsupervised learning

- Discovering graph structure
- Dimensionality reduction
- Latent factors revealing
- Image inpainting
- Collaborative filtering
- Market basket analysis

Open-problems in machine learning

- Parametric vs. non-parametric models
- The curse of dimensionality
- Overfitting and underfitting
- Model selection
- No free lunch theorem

All models are wrong, but some models are useful *– there is no single best model that works optimally or all kind of data* (***George Box***, 1987)

System identification

System identification – the discipline in statistics and machine learning that *builds mathematical models* of *dynamical systems* from measured data. It also includes the *optimal design of experiments* for efficiently generating informative data *for fitting* such models *as well as model reduction*.

System identification models

- ***White-box model*** – the mathematical model of system is known in advanced. But in many cases such model will be overly complex and possibly even impossible to obtain in reasonable time due to the complex nature of many systems and processes;

- ***Grey-box model*** – a certain model based on both insight into the system and experimental data is constructed. This model does however still have a number of unknown free parameters which can be estimated using system identification;

- ***Black-box model*** – No prior model is available. Most system identification algorithms are of this type.

Conclusion

- The main types of machine learning task were considered;
- Practical applications and challenges in machine learning domain were shown;
- The problem of system identification as machine learning task is established. Key types of system models were considered.