EPFL

PHYS-467   **Machine learning for physicists**

Zdeborová Lenka

Cursus	Sem.	Type
Chimiste	MA1, MA3	Opt.
Ing.-phys	MA1, MA3	Opt.
Physicien	MA1, MA3	Opt.
Quantum Science and Engineering	MA1, MA3	Opt.

Language of teaching	English
Credits	4
Session	Winter
Semester	Fall
Exam	Written
Workload	120h
Weeks	14
Hours	**4 weekly**
Lecture	2 weekly
Exercises	2 weekly
Number of positions	

Summary

Machine learning and data analysis are becoming increasingly central in sciences including physics. In this course, fundamental principles and methods of machine learning will be introduced and practised.

Content

* Examples and types of problems that machine learning can solve.
* Linear regression in matrix notation. The concept of prediction, estimation. Least squares method. High-dimensional underdetermined problems and the concept of regularization aka ridge regression. Polynomial regression. The concept of bias and variance trade-off and overfitting. Usage of train, validation and test sets.
* Reminder of key concept from probability theory. Bayesian inference, maximum likelihood and maximum a posteriori estimation.
* Least-squares as maximum likelihood of probabilistic model with additive Gaussian noise. Regularization as a prior. Relation with inverse problems in signal processing. Generalized linear model.
* Robust regression, sparse regression, LASSO. Role of sparsity for variable selection. Compressed sensing.
* Gradient descent and stochastic gradient descent.
* Linear classification. Examples of classification losses. Logistic regression and its probabilistic interpretation. Multi-class classification, one-hot-encoding of classes, cross-entropy loss.Data that are not linearly separable. K-nearest neighbours. Curse of dimensionality.
* Unsupervised learning, dimensionality reduction, low-rank approximation. Singular values decomposition (SVD) and principal component analysis (PCA).
Examples: Recommender systems, reconstruction of Europe geography from human genome, spin-glass card game aka planted spin glass model.
* Analogy between learning/inference in high dimension and statistical mechanics. Maximum a posteriori estimation as search for the ground state.
Minimum mean squared error estimator. Bayesian inference as sampling from the Boltzmann measure.
* Monte Carlo Markov Chains and their basic principles. Metropolis-Hastings update rule.
Gibbs sampling aka heat bath. Simulated annealing.
Bayesian learning of hyper-parameters, expectation maximization algorithm.
* Clustering. The k-means algorithm. Gaussian mixture model.
* Data generative models. Boltzmann machine. Maximum entropy principle. Training algorithm for the Boltzmann machine.
* Non-linear regression as linear regression in feature space. Representor theorem. Kernel methods as infinite-dimensional feature spaces. Kernel ridge regression. Examples of kernels and their feature spaces. Kernels as universal approximations. Classification with kernels, support vector machines.
* Random feature regression as approximation of kernels. One hidden-layer neural networks as features learning machines. Neural networks as universal approximators. Worst case computational hardness of training. Multi-layer neural networks as learning features of features.
* Deep learning for regression and classifications. Terminology of multi-layer feed-forward neural networks. Training with stochastic gradients descent aka the back-propagation algorithm. Discussion of hyper-parameters to be set when using neural networks. Historical notes and comments on performance of deep learning.
Importance of locality and translational symmetry. Convolutional neural networks for image classification. Design and

terminology of convolutional and pooling layers.

Modus operandi of deep neural networks. Over-parametrization and lack of overfitting. Double descent behaviour replaced the bias-variance trade-off. Interpolation of the training set and its consequences for training, implicit regularization.

Concept of transfer learning, adversarial examples, data augmentation.

* Unsupervised learning using tricks from deep learning. Boltzmann machine with hidden variables, restricted Boltzmann machines, contrastive divergence algorithm.

The principle of auto-encoder, its training and usage. Principles of generative adversarial networks (GANs), their usage and training.

Learning Prerequisites

Important concepts to start the course

Basic notions in probability, analysis and basic familiarity with programming. Some notions of statistical physics will be used to support this lecture.

Learning Outcomes

By the end of the course, the student must be able to:
- Use basic tools for data analysis and for learning from data
- Explain basic principles of data analysis and learning from data
- List and explain machine learning tools suited for a given problem.

Teaching methods

2h of lecture + 2h of excercise (exercise mostly with a computer)

Assessment methods

Final written exam counting for 50% and several graded homeworks during the semester counting for the other 50%.

Resources

Bibliography

A high-bias, low-variance introduction to Machine Learning for physicists. Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, David J. Schwab, https://arxiv.org/abs/1803.08823.

Text book "Information Theory, Inference, and Learning Algorithms" by David MacKay.
Polycopie of the lecture available in Moodle.

Ressources en bibliothèque

- A high-bias, low-variance introduction to Machine Learning for physicists. Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, David J. Schwab
- Information Theory, Inference, and Learning Algorithms /David MacKay

Moodle Link

- https://go.epfl.ch/PHYS-467