Machine Learning for Particle Physics

Monday 6 June 2016 - Friday 10 June 2016

CERN

Scientific Programme

Overview of the course. Brief introduction to various concepts in machine learning. Supervised and unsupervised learning. Classification versus regression. Statistical learning and Bayes Rule. The learning curve and overtraining. Model bias and variance. Cross-validation. Variable transformation and feature learning. Supervised feature ranking and selection.

Variable transformations. Principal component analysis. Multidimensional scaling. Kernel PCA. Feature learning by autoencoder or a similar algorithm.

Classification:

Decision trees. CART and C4.5. Splitting rules and stopping criteria. Accounting for variable interactions. Strengths and weaknesses of decision trees. Random forest.

Support vector machines. Linear SVM for separable classes. Linear SVM for inseparable classes. Non-linear SVM. Computational issues. Multiclass SVM by reduction to binary.

Feature ranking and selection. Estimation of feature importance by neighborhood component analysis and random forest. Sequential forward addition and backward elimination. Use of statistical tests for comparing accuracies of two sets of features.