

## Machine Learning

**Objective:** Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data.

The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns. However, instead of extracting data for human comprehension -- as is the case in data mining applications -- machine learning uses that data to detect patterns in data and adjust program actions accordingly. Example: Facebook's News Feed uses machine learning to personalize each member's feed.

### Topics to be covered in Workshop (20% Theory & 80% Hands-On Session)

1. **Introduction:** Definition of learning systems. Goals and applications of machine learning. Aspects of developing a learning system: training data, concept representation, function approximation.
2. **Inductive Classification:** The concept learning task. Concept learning as search through a hypothesis space. General-to-specific ordering of hypotheses.
3. **Decision Tree Learning:** Representing concepts as decision trees. Recursive induction of decision trees. Picking the best splitting attribute: entropy and information gain. Searching for simple trees and computational complexity. Occam's razor. Over fitting, noisy data, and pruning.
4. **Ensemble Learning:** Using committees of multiple hypotheses. Bagging, boosting, and DECORATE. Active learning with ensembles.

5. **Experimental Evaluation of Learning Algorithms:** Measuring the accuracy of learned hypotheses. Comparing learning algorithms: cross-validation, learning curves, and statistical hypothesis testing.
6. **Computational Learning Theory:** Models of learnability: learning in the limit; probably approximately correct (PAC) learning. Sample complexity: quantifying the number of examples needed to PAC learn.
7. **Rule Learning:** Propositional and First-Order: Translating decision trees into rules. Heuristic rule induction using separate and conquer and information gain.
8. **Artificial Neural Networks:** Neurons and biological motivation. Linear threshold units. Perceptron's: representational limitation and gradient descent training.
9. **Support Vector Machines:** Maximum margin linear separators. Quadratic programming solution to finding maximum margin separators. Kernels for learning non-linear functions.
10. **Bayesian Learning:** Probability theory and Bayes rule. Naive Bayes learning algorithm.
11. **Instance-Based Learning:** Constructing explicit generalizations versus comparing to past specific examples. K-Nearest-neighbour algorithm. Case-based learning.
12. **Text Classification:** Bag of words representation. Vector space model and cosine similarity. Relevance feedback and Rocchio algorithm.
13. **Clustering and Unsupervised Learning:** Learning from unclassified data. Clustering. Hierarchical Agglomerative Clustering. K-means partitional clustering.
14. **Language Learning:** Classification problems in language: word-sense disambiguation, sequence labeling. Hidden Markov models (HMM's).