Applied Machine Learning

Data Science and Analytics Institute
The University of Oklahoma

Overview

Today, machine learning (ML) permeates our lives to an extent that could only be imagined a few short years ago. Its applications are all around us. From empowering personal assistant tools and fraud detection software to finding cures for disease years earlier than expected, ML has and is transforming businesses, governments, and society.

This training provides an introduction to machine learning, including overviews of data science, problem types and data science and machine learning tools. Concepts of exploratory data analysis, supervised learning, unsupervised learning and connectionist components are further explored. The training will conclude with a demonstration of select hands-on applications.

Course Objectives

1. Provide an overview of machine learning and related tools and topics.
2. Facilitate the dissemination of knowledge and the use of machine learning throughout the organization.
3. Increase participants' workplace effectiveness with machine learning applications.

Learning Outcomes

After completing the course, participants will be able to:

1. Identify and recall ML and intelligent analytics methods to discover knowledge and insight from data.
2. Apply exploratory analysis, supervised and unsupervised learning, and dimensionality reduction techniques towards industry use cases.
3. Given appropriate datasets and tools, use intelligent analytics methods to produce insight or solutions to a problem.

Intended Audience

Course participants should be familiar with computer programming and statistics.

Certification

Participants who register and complete the course will earn a Certificate of Participation and Completion from the University of Oklahoma.

Topic Overview

Day 1

Introduction: welcome and overview

- Welcome: introductions
- Welcome: expectations
- Welcome: goals and feedback
- Welcome: software overview

Brief Machine Learning Overview

- Definition
- Relationship to artificial intelligence, machine learning, deep learning, data science
- Applications: why use machine learning?
- Applications: real-world examples
- Overview: problem types (SL, UL, RL)
- Overview: batch vs. online learning
- Overview: Instance- vs. model-based
- Overview: trends (R&D, technical performance, economy & education)

How Supervised Learning Works

- Gathering data: spam detection or house price prediction
- Data pre-processing: feature engineering
- Build a model: SVM or linear regression

Challenges in Machine Learning

- Insufficient quantity of data
- Non-representative training data
- Poor-quality data
- Irrelevant features
- Underfitting
- Overfitting

Generalization: definition, validation techniques

End-to-End Machine Learning Example

- Organizing framework: business understanding, data preparation, data modeling, business validation
- Hands-on example in Colab

Day 2

Review

- Summary of Day 1
- Call for questions / select exercise solutions

Basic practice

- Organizing framework revisited
- Exploratory analysis
- Data preparation: feature engineering
- Fundamental algorithms: linear regression, logistic regression, decision tree learning, support vector machines, kNN
- Anatomy of a learning algorithm: building blocks (loss function, optimization criterion, optimization routine), gradient descent

Assessing performance

- Regression: RMSE
- Classification: confusion matrix, precision/recall, accuracy, AUC

Improving performance

- Hyperparameter tuning
- Regularization

Learning algorithm selection: guiding questions and algorithm peculiarities

Day 3

Neural Networks and Deep Learning

- Neural networks: MLPs, Feed-Forward Neural Network architecture
- Deep learning: CNN, RNN

Ensemble Methods

- Boosting/bagging (RF, Gradient Boosting)
- Averaging, majority vote, stacking

Unsupervised Learning

- Density estimation
- Clustering: k-Means, DBSCAN and HDBSCAN, others
- Dimensionality reduction: PCA, UMAP