Machine Learning using Python

Description

This Machine Learning with Python course dives into the basics of machine learning using an approachable, and well-known, programming language. l learn about Supervised vs. Unsupervised Learning, look into how Statistical Modeling relates to Machine Learning, and do a comparison of each.

Expectations and Goals

This course helps participants understand what data scientists do, the problems they solve, and the tools and techniques they use. Through in-class simulations, participants apply data science methods to real-world challenges in different industries and, ultimately, prepare for data scientist roles in the field.

Prerequisites

This course is suitable for students, developers, data analysts, and statisticians with basic knowledge of Computer Science and python programming.

Course Schedule

Module	Topic
Module 1	The Machine Learning Landscape What Is Machine Learning? Why Use Machine Learning? Types of Machine Learning Systems Supervised/Unsupervised Learning Batch and Online Learning Instance-Based Versus Model-Based Learning Main Challenges of Machine Learning Insufficient Quantity of Training Data Non representative Training Data Poor-Quality Data Irrelevant Features Overfitting the Training Data Underfitting the Training Data Testing and Validating
Module 2	NumPy Basics: Arrays and Vectorized Computation The NumPy ndarray: A Multidimensional Array Object Creating ndarrays Data Types for ndarrays Arithmetic with NumPy Arrays Basic Indexing and Slicing Boolean Indexing Fancy Indexing Transposing Arrays and Swapping Axes Universal Functions: Fast Element-Wise Array Functions Array-Oriented Programming with Arrays Expressing Conditional Logic as Array Operations Mathematical and Statistical Methods Methods for Boolean Arrays

	Sorting Unique and Other Set Logic File Input and Output with Arrays Linear Algebra Pseudorandom Number Generation Example: Random Walks Simulating Many Random Walks at Once
Module 3	Getting Started with pandas Introduction to pandas Data Structures Series DataFrame Index Objects Essential Functionality Reindexing Dropping Entries from an Axix Indexing, Selection, and Filtering Integer Indexes Arithmetic and Data Alignment Function Application and Mapping Sorting and Ranking Axis Indexes with Duplicate Labels Summarizing and Computing Descriptive Statistics Correlation and Covariance Unique Values, Value Counts, and Membership
Module 4	End-to-End Machine Learning Project Working with Real Data Look at the Big Picture Frame the Problem Select a Performance Measure Check the Assumptions Get the Data Create the Workspace Download the Data Take a Quick Look at the Data Structure Create a Test Set Discover and Visualize the Data to Gain Insights Visualizing Geographical Data Looking for Correlations Experimenting with Attribute Combinations Prepare the Data for Machine Learning Algorithms Data Cleaning Handling Text and Categorical Attributes Custom Transformers Feature Scaling Transformation Pipelines

	Select and Train a Model Training and Evaluating on the Training Set Better Evaluation Using Cross-Validation Fine-Tune Your Model Grid Search Randomized Search Ensemble Methods Analyze the Best Models and Their Errors Evaluate Your System on the Test Set Launch, Monitor, and Maintain Your System
Module 5	Classification MNIST Training a Binary Classifier Performance Measures Measuring Accuracy Using Cross-Validation Confusion Matrix Precision and Recall Precision/Recall Tradeoff The ROC Curve Multiclass Classification Error Analysis Multilabel Classification Multioutput Classification
Module 6	Training Models Linear Regression The Normal Equation Computational Complexity Gradient Descent Batch Gradient Descent Stochastic Gradient Descent Mini-batch Gradient Descent Polynomial Regression Learning Curves Regularized Linear Models Ridge Regression Lasso Regression Elastic Net Early Stopping Logistic Regression Estimating Probabilities Training and Cost Function Decision Boundaries Softmax Regression
Module 7	Support Vector Machines Linear SVM Classification

	Soft Margin Classification Nonlinear SVM Classification Polynomial Kernel Adding Similarity Features Gaussian RBF Kernel Computational Complexity SVM Regression Under the Hood Decision Function and Predictions Training Objective Quadratic Programming The Dual Problem, Generalized Lagrangian for the Hard margin problem and KKT multiplier Kernelized SVM and Mercer's Theorem Online SVMs and Hinge loss and how is it used in SGD classification
Module 8	Decision Tree Training and visualizing a decision tree Making predictions Estimating class probabilities The CART (Classification and Regression Tree) training algorithm Computational Complexity Gini Impurity or Entropy? Regularization hyper parameters Regression Instability
Module 9	Ensemble Learning and Random Forest Voting Classifiers Bagging and Pasting Random Patches and Random subspaces Random Forests Boosting Stacking
Module 10	Dimensionality Reduction The Curse of Dimensionality Main Approaches for Dimensionality Reduction Projection Manifold Learning PCA Preserving the Variance Principal Components Projecting Down to d Dimensions Using Scikit-Learn Explained Variance Ratio Choosing the Right Number of Dimensions PCA for Compression Randomized PCA

	Incremental PCA Kernel PCA Selecting a Kernel and Tuning Hyperparameters LLE Other Dimensionality Reduction Techniques
Module 11	Unsupervised Learning Techniques Clustering K-Means Limits of K-Means Using Clustering for Image Segmentation Using Clustering for Preprocessing Using Clustering for Semi-Supervised Learning DBSCAN Other Clustering Algorithms Gaussian Mixtures Anomaly Detection Using Gaussian Mixtures Selecting the Number of Clusters Bayesian Gaussian Mixture Models Other Algorithms for Anomaly and Novelty Detection
Module 12	Project work and documentation