Course Outline

Practical Data Science with Amazon SageMaker

Course Summary

Description

In this intermediate-level course, individuals learn how to solve a real-world use case with Machine Learning (ML) and produce actionable results using Amazon SageMaker. This course walks through the stages of a typical data science process for Machine Learning from analyzing and visualizing a dataset to preparing the data, and feature engineering. Individuals will also learn practical aspects of model building, training, tuning, and deployment with Amazon SageMaker. Real life use cases include customer retention analysis to inform customer loyalty programs.

Objectives

At the end of this course, students will be able to:

- Prepare a dataset for training
- Train and evaluate a Machine Learning model
- Automatically tune a Machine Learning model
- Prepare a Machine Learning model for production
- Think critically about Machine Learning model results

Topics

- Introduction to Machine Learning
- Introduction to Data Prep and SageMaker
- Problem formulation and Dataset Preparation
- Data Analysis and Visualization
- Training and Evaluating a Model
- Automatically Tune a Model
- Deployment / Production Readiness
- Relative Cost of Errors

Audience

This course is designed for Developers and Data Scientists.

Prerequisites

Before taking this course, students should have:

- Familiarity with Python programming language
- Basic understanding of Machine Learning
- Basic understanding of AWS Cloud infrastructure (Amazon S3 and Amazon CloudWatch)

Duration

One day

Course Outline

Practical Data Science with Amazon SageMaker

Course Outline

I. Introduction to Machine Learning
- A. Types of ML
- B. Job Roles in ML
- C. Steps in the ML pipeline

II. Introduction to Data Prep and SageMaker
- A. Training and Test dataset defined
- B. Introduction to SageMaker
- C. Demo: SageMaker console
- D. Demo: Launching a Jupyter notebook

III. Problem formulation and Dataset Preparation
- A. Business Challenge: Customer churn
- B. Review Customer churn dataset

IV. Data Analysis and Visualization
- A. Demo: Loading and Visualizing your dataset
- B. Exercise 1: Relating features to target variables
- C. Exercise 2: Relationships between attributes
- D. Demo: Cleaning the data

V. Training and Evaluating a Model
- A. Types of Algorithms
- B. XGBoost and SageMaker
- C. Demo 5: Training the data
- D. Exercise 3: Finishing the Estimator definition
- E. Exercise 4: Setting hyperparameters
- F. Exercise 5: Deploying the model
- G. Demo: Hyperparameter tuning with SageMaker
- H. Demo: Evaluating Model Performance

VI. Automatically Tune a Model
- A. Automatic hyperparameter tuning with SageMaker
- B. Exercises 6-9: Tuning Jobs

VII. Deployment / Production Readiness
- A. Deploying a model to an endpoint
- B. A/B deployment for testing
- C. Auto Scaling Scaling
- D. Demo: Configure and Test Autoscaling
- E. Demo: Check Hyperparameter tuning job
- F. Demo: AWS Autoscaling
- G. Exercise 10-11: Set up AWS Autoscaling

VIII. Relative Cost of Errors