55375A: Fundamentals of Machine Learning

Course Details

Course Code: 55375A

Duration: 5 days

Notes:

- This course syllabus should be used to determine whether the course is appropriate for the students, based on their current skills and technical training needs.
- Course content, prices, and availability are subject to change without notice.
- Terms and Conditions apply

Academy IT Pty Ltd
Level 4, 45 Grenfell Street
ADELAIDE 5000

Email: sales@academyit.com.au
Web: www.academyit.com.au

Phone: 08 7324 9800
Brian: 0400 112 083

Elements of this syllabus are subject to change.

About this course

This five-day instructor-led course is intended for IT professionals who wish to learn the fundamental principles of machine learning classification models. Students are typically developers or data analysts that have some programming skills, with a mathematical background. In this course, students will learn about classification machine learning models together with the algorithms that underpin these models. Students will also learn about the need for data quality, and how the correct selection of data can help to build more accurate models. Students will also learn how to test and validate models.

This course is generic, and doesn't depend on any particular platform (Azure, AWS, Google Cloud etc), but does require some basic familiarity with Python. Students should also have some basic understanding of differential calculus, probability, and statistics.

Note: This course focuses on classification models. Two further courses cover cluster models and regression models. This course does not cover deep learning or neural networks, although the elements of this course can provide a foundation for further study in those areas.

Audience Profile

This course is intended for developers and analysts who are new to machine learning, want to understand how machine learning classification models work, and need to understand how to build high-quality models. Attendees must have some familiarity with Python, an understanding of matrix and vector arithmetic and some basic familiarity with probability, statistics, and differential calculus.

At Course Completion

After completing this course, students will be able to:

- Explain the uses of different types of machine learning models (classification, clustering, and regression).
- Understand common machine learning classification algorithms.
- Create and test a classification machine learning model.
- Create non-binary classification models.
- Summarize the statistical concepts utilized by many machine learning models.
- Select features for a classification model
- Measure the performance of a classification model
- Build classification models that can handle imbalanced datasets

Prerequisites

- Familiarity with the Python programming language.
- An understanding of basic probability and statistics.
- Ideally, some familiarity with the basics of differential calculus.

Module 1: Introduction to Machine Learning Models

This module introduces machine learning together with the classification, clustering, and regression machine learning models. Students will learn the purpose of these models, and the types of problems to which students can apply them.

Lessons

- Lesson 1: Understanding Machine Learning
- Lesson 2: Understanding Machine Learning Models
- Lesson 3: Understanding the Process for Creating a Machine Learning Model
- Lesson 4: Reviewing Essential Math Concepts
- Lesson 5: Using Common Python Libraries and Packages for Machine Learning

Lab 1: Introduction to Machine Learning Models

- Examine several scenarios, and determine the most appropriate type of machine learning model for each scenario.

After completing this module, students will be able to:

- Describe the purpose of machine learning.
- Explain different types of machine learning models, and when to use each type.
- Describe how to perform common differential calculus operations, and the use of natural logs.
- Describe common Python libraries and packages that support building and testing machine learning models

Module 2: Understanding Classification Algorithms

This module describes different algorithms that are commonly used to create a classification machine learning model. It provides a tour through the algorithms, summarizing their strengths and weaknesses, and when each is most appropriate.

Lessons

- Lesson 1: Understanding Decision Trees
- Lesson 2: Understanding Random Forests
- Lesson 3: Understanding Gradient Boosted Trees
- Lesson 4: Understanding XGBoost
- Lesson 5: Understanding Logistic Regression
- Lesson 6: Understanding the K-Nearest Neighbors Algorithm
- Lesson 7: What are Other Common Algorithms?

After completing this module, students will be able to:

- Describe the Decision Tree algorithm.
- Explain how Random Forests work.
- Understand how trees can use Gradient Boosting to make more accurate predictions.
- Describe how Extreme Gradient Boosting extends the Gradient Boosting algorithm
- Explain Logistic Regression.
- Describe the K-Nearest Neighbors algorithm.
- Summarize the Support Vector Machine, Linear Discriminant Analysis, and the Naïve Bayes algorithms.

Module 3: Creating a Classification Model

This module provides an overview of the essential steps in building a machine learning model: data preparation, model construction and tuning, and testing and validation.

Lessons

- Lesson 1: Preparing the Data
- Lesson 2: Building and Fitting a Model
- Lesson 3: Testing and Validating a Classification Model

Lab 1: Creating a Classification Machine Learning Model

- Examine the problem and the data
- Import and prepare the data
- Use a decision tree to classify the data
- Test and validate the results

- Use a random forest to classify the data, and compare the performance and results
- Use a gradient boosted tree to classify the data, and compare the performance and results
- Use Logistic Regression to classify the data, and compare the performance and results
- Use the K-Nearest Neighbours algorithm to classify the data, and compare the performance and results

After completing this module, students will be able to:

- Prepare a dataset for classification
- Create a classification machine learning model using the Scikit-Learn Python library
- Test and validate a classification machine learning model

Module 4: Understanding Binary and Non-Binary Classification

This module describes the differences between binary and multi-valued classification and shows how to ceate a multi-class classification model.

Lessons

- Lesson 1: Understanding Multi-class Classification
- Lesson 2: Understanding the One versus Rest and One versus One Algorithms
- Lesson 3: Understanding Multi-label Classification

Lab 1: Creating a Multi-class and Multi-lavel Machine Learning Model

- Examine the problem and the data (Mobile Phone Prices – grouping phones into different price categories depending on their features)
- Import and prepare the data
- Use a decision tree to classify the data
- Test and validate the results
- Use Logistic Regression to classify the data, and compare the performance and results
- Use the K-Nearest Neighbours algorithm to classify the data, and compare the performance and results
- Use Gradient Boosting to create a multi-label model that identifies whether a phone should have 4G and WiFi capabilities based on its other features and price category.

After completing this module, students will be able to:

- Describe multi-class classification and build models that support multi-class classification
- Explain the One versus Rest and One versus One algorithms for implementing multi-class classification
- Describe multi-label classification and how it differs from multi-class classification

Module 5: Reviewing Statistics Concepts

This module summarizes key statistics terminology, and some common techniques used to analyze the distribution, scale, and relationships between items in a dataset. This information is essential to understanding the validity of a machine learning model.

Lessons

- Lesson 1: Understanding Statistical Sampling
- Lesson 2: Understanding Measures of Central Tendency
- Lesson 3: Calculating Measures of Dispersion
- Lesson 4: Evaluating the Sampling Strategy
- Lesson 5: Estimating Confidence Intervals and Sampling Error
- Lesson 6: Quantifying the Differences between Data Distributions

After completing this module, students will be able to:

- Describe the importance of statistical sampling for building an accurate machine learning model.
- Explain the measures of central tendency used to describe statistical models.
- Use measures of dispersion to understand how the data in a sample varies from the centre.
- Evaluate a sampling strategy

- Assess confidence intervals and sampling error and measure statistical significance.
- Quantify the differences between data distributions.

Module 6: Exploring Data and Selecting Features and Algorithms

This module explains how to refine a machine learning model, by selecting the most relevant features from the dataset, examining the distribution of values, investigating correlation between features, normalizing data, and removing bias. This is useful in refining the features of the dataset used to create a machine learning model.

Lessons

- Lesson 1: Graphing Data to Examine Relationships and Identify Skew
- Lesson 2: What is Correlation and Casuality?
- Lesson 3: Selecting Model Features
- Lesson 4: Extracting and Scaling Features
- Lesson 5: Creating a pre-processing pipeline

After completing this module, students will be able to:

- Create graphs and charts to explore a dataset and identify skew.
- Explain how to detect correlation between the variables in a dataset.
- Select the features of a dataset that have the greatest impact on the predictions it makes.
- Combing and scale features in dataset.
- Use pipelines to automate data pre-processing and model construction.

Module 7: Measuring the Performance of a Classification Model

This module describes how to assess the accuracy and performance for a classification model, and how to balance precision and recall where appropriate.

Lessons

- Lesson 1: Understanding Performance Measures for a Classification Model
- Lesson 2: Understanding Regularization to Reduce Overfitting
- Lesson 3: Evaluating a Model

Lab 1: Refining a Machine Learning Model

- Examine the scenario (sample data set with many irrelevant dependent, and/or badly scaled features with non-standard distributions)
- Build a Logistic Regression model to classify the data without any modifications to the data
- Examine the results and measure the performance
- Explore and refine the dataset
- Recreate and retest the model
- Repeat until the performance is optimized
- Compare the performance of two models constructed using different algorithms.

After completing this module, students will be able to:

- Describe performance metrics for classification models.
- Explain how to use regularization to prevent overfitting.
- Evaluate and compare classification models.

Module 8: Understanding Imbalanced Classification

This lesson discusses the problems that can arise when using an imbalanced dataset to create a classification model, how to recognize potential problems, and how to address them.

Lessons

- Lesson 1: Understanding Imbalanced Classification
- Lesson 2: Calibrating a Model
- Lesson 3: Using Data Sampling to Balance a Dataset
- Lesson 4: Understanding Evaluation Metrics for an Imbalanced Dataset

Lab 1: Handling Imbalanced Data

- Classifying Imbalanced Data
- Start with an imbalanced dataset and build a vanilla Random Forest model.
- Generate the ROC and accuracy, and evaluate the results (they are not as good as they first appear).
- Balance the dataset [sampling] and run the model again.

- Evaluate the model again.
- Build another model with the original imbalanced dataset, and then calibrate and evaluate the model
- Combine models using the VotingClassifier class and evaluate the result

After completing this module, students will be able to:

- Identify imbalance in a dataset.
- Calibrate a machine learning model to mitigate the effects of imbalance in a dataset.
- Adjust the distribution of data in an imbalanced dataset by using sampling.
- Measure the effects of imbalanced data on a machine learning model.