Data Scientist Syllabus

Contact Info

While going through the program, if you have questions about anything, you can reach us at support@udacity.com. For help from Udacity Mentors and your peers visit the Udacity Classroom.

Nanodegree Program Info

Version: 2.0.0

Length of Program: 81 Days*

** This is a self-paced program and the length is an estimation of total hours the average student may take to complete all required coursework, including lecture and project time. Actual hours may vary.*

Part 1: Introduction

Part 2: Supervised Learning

Project: Finding Donors for CharityML

You've covered a wide variety of methods for performing supervised learning -- now it's time to put those into action!

Supporting Lessons

Lesson	Summary
Linear Regression	Linear regression is a very effective algorithm to predict numerical data.
Perceptron Algorithm	The perceptron algorithm is an algorithm for classifying data. It is the building block of neural networks.
Decision Trees	Decision trees are a structure for decision-making where each decision leads to a set of consequences or additional decisions.
Naive Bayes	Naive Bayesian Algorithms are powerful tools for creating classifiers for incoming labeled data.
Support Vector Machines	Support vector machines are very effective models used for classification.
Ensemble Methods	Bagging and boosting are two common ensemble methods for improving the accuracy of supervised learning approaches.
Supervised Learning Assessment	Test your Supervised Learning concepts with a quick assessment.

Part 3: Unsupervised Learning

Project: Creating Customer Segments

Now that you've learned a lot about unsupervised learning, it's time to apply that to a project.

Supporting Lessons

Lesson	Summary
Clustering	Clustering is one of the most common methods of unsupervised learning. Here, we'll discuss the K-means clustering algorithm.
Clustering Mini-Project	In this mini-project, you will use K-means to cluster movie ratings and use those clusters to provide movie recommendations.
Hierarchical and Density-based Clustering	We continue to look at clustering methods. Here, we'll discuss hierarchical clustering and density-based clustering (DBSCAN).
Gaussian Mixture Models and Cluster Validation	In this lesson, we discuss Gaussian mixture model clustering. We then talk about the cluster analysis process and how to validate clustering results.
Feature Scaling	Feature scaling is an important pre-processing step when performing unsupervised learning to allow multiple features to be analyzed together.
PCA	PCA, principal component analysis, is a method for feature selection that turns a set of correlated variables into the underlying set of orthogonal variables.
PCA Mini-Project	In this mini-project, you'll apply principal component analysis to facial recognition.
Random Projection and ICA	In this lesson, we will look at two methods for feature extraction and dimensionality reduction: Random Projection and Independent Component Analysis (ICA)
Unsupervised Learning Assessment	Test your understanding of unsupervised learning with a quick assessment.

Project: Improve Your LinkedIn Profile

Find your next job or connect with industry peers on LinkedIn. Ensure your profile attracts relevant leads that will grow your professional network.

Project: Optimize Your GitHub Profile

Other professionals are collaborating on GitHub and growing their network. Submit your profile to ensure your profile is on par with leaders in your field.

Part 4: Big Data & Map Reduce

Project: Explore and Summarize Data

Choose one of Udacity's curated datasets or find one of your own and perform a complete exploratory data analysis on the data using R.

Udacity

Generated Mon May 11 14:10:27 PDT 2020