Data Scientist Nanodegree 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: 1.0.0

Length of Program: 213 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: Welcome to the Nanodegree

Learn about what you will build, meet your instructors, and learn how to be successful in the Data Scientist Nanodegree Program!

Part 2: Supervised Learning

Learn to build supervised machine learning models to make data-informed decisions. Learn to evaluate and validate the quality of your models.

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
Machine Learning Bird's Eye View	Before diving into the many algorithms of machine learning, it is important to take a step back and understand the big picture associated with the entire field.
Linear Regression	Linear regression is one of the most fundamental algorithms in machine learning. In this lesson, learn how linear regression works!
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. Specifically Naive Bayes is frequently used with text data and classification problems.
Support Vector Machines	Support vector machines are a common method used for classification problems. They have been proven effective using what is known as the 'kernel' trick!
Ensemble Methods	Bagging and boosting are two common ensemble methods for combining simple algorithms to make more advanced models that work better than the simple algorithms would on their own.
Model Evaluation Metrics	Learn the main metrics to evaluate models, such as accuracy, precision, recall, and more!
Training and Tuning	Learn the main types of errors that can occur during training, and several methods to deal with them and optimize your machine learning models.

Part 3: Deep Learning

Gain a solid foundation in neural networks, deep learning, and PyTorch.

Project: Create Your Own Image Classifier

In this project, you'll build a Python application that can train an image classifier on a dataset, then predict new images using the trained model.

Supporting Lessons

Lesson	Summary
Introduction to Neural Networks	In this lesson, Luis will give you solid foundations on deep learning and neural networks. You'll also implement gradient descent and backpropagation in python right here in the classroom.
Implementing Gradient Descent	Mat will introduce you to a different error function and guide you through implementing gradient descent using numpy matrix multiplication.
Training Neural Networks	Now that you know what neural networks are, in this lesson you will learn several techniques to improve their training.
Keras	In this section you'll get a hands-on introduction to Keras. You'll learn to apply it to analyze movie reviews.
Deep Learning with PyTorch (updated version)	Learn how to use PyTorch for building deep learning models.
Deep Learning with PyTorch	Learn how to use PyTorch for building deep learning models

Part 4: Unsupervised Learning

Learn to build unsupervised machine learning models, and use essential data processing techniques like scaling and PCA.

Project: Creating Customer Segments with Arvato

In this project, you'll apply your unsupervised learning skills to two demographics datasets, to identify segments and clusters in the population, and see how customers of a company map to them.

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.
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.
Dimensionality Reduction and PCA	Often we need to reduce a large number of features in our data to a smaller, more relevant set. Principal Component Analysis, or PCA, is a method of feature extraction and dimensionality reduction.
Random Projection and ICA	In this lesson, we will look at two other methods for feature extraction and dimensionality reduction: Random Projection and Independent Component Analysis (ICA).

Part 5: Congratulations

Congratulations! You have completed Term 1 of DSND! Now what?

Part 6: Welcome to Term 2

In this term, you'll master the skills necessary to become a successful Data Scientist. You'll work on projects designed by industry experts, and learn to run data pipelines, design experiments, build recommendation systems, and deploy solutions to the cloud.

Part 7: Introduction to Data Science

Learn the data science process, including how to build effective data visualizations, and how to communicate with various stakeholders.

Project: Write A Data Science Blog Post

Put your skills to work! Use the CRISP-DM process to create a blog and Github post to show off your skills.

Supporting Lessons

Lesson	Summary
The Data Science Process	In this lesson, you will learn about CRISP-DM and how you can apply it to many data science problems.
Communicating to Stakeholders	In this lesson, you will be creating a post to communicate your findings via Medium.

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 8: Software Engineering

Software engineering skills are increasingly important for data scientists. In this course, you'll learn best practices for writing software. Then you'll work on your software skills by coding a Python package and a web data dashboard.

Part 9: Data Engineering

In data engineering for data scientists, you will practice building ETL, NLP, and machine learning pipelines. This will prepare you for the project with our industry partner Figure 8.

Project: Disaster Response Pipelines

You'll build a machine learning pipeline to categorize emergency messages based on the needs communicated by the sender.

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.

Part 10: Experimental Design & Recommendations

Learn to design experiments and analyze A/B test results. Explore approaches for building recommendation systems.

Project: Recommendations with IBM

Put your skills to work to make recommendations for IBM Watson Studio's data platform.

Part 11: Data Scientist Capstone

Leverage what you've learned throughout the program to build your own open-ended Data Science project. This project will serve as a demonstration of your valuable abilities as a Data Scientist.

Project: Capstone Project

Now you will put your Data Science skills to the test by solving a real world problem using all that you have learned throughout the program.

Part 12: Congratulations

Congratulations on your completion of the Data Scientist Nanodegree!

Udacity

Generated Wed Oct 23 14:55:11 PDT 2019