Dive Deeper

In the previous module, you learned about different data transformation techniques, and particularly about pre-processing transformations and scalers that you should apply early on in the data analysis process to whip your data into shape. In this module, you touched on unsupervised clustering. Then you learned about splitting your data between a testing and a training set, so that you don't mistakenly inflate your supervised learning algorithm's scores, by giving them good marks for correctly identifying samples you've already given them the answer to. After that, you moved into your first classification algorithm, K-Nearest Neighbors, one of the algorithm many new and budding data scientists choose to program for themselves as a start to writing their own classification algorithms. And then lastly, you took the plunge with linear regression and learned how to predict continuous outputs.

Congratulations on making it this far! This module is without a doubt, the longest module in this course. You're now well over the hump and all that's left to do is wrap up! In the next module, you will learn about three very powerful and distinct classification algorithms, including one new one that was invented rather recently. Before you get there, take a moment to dive deeper into the extra learning material we've included for you below!

K-Means

- +30 Methods: On Determining the Number of Clusters
- Why K-Means Isn't Great For Sparse Datasets
- K-Means In Sparse High-D Datasets

- K-Medoids
- Initial Centroid Positioning

K-Neighbors

- Implementing K-Neighbors From Scratch

CDR

- Call Detail Records Generator
- Another Call Detail Records Generator
- Sample (Real) CDR Dataset
- CDR Data Applied in the Wild
- KMeans + CDR

Oddities

- When to Use Feature Scaling
- SciKit-Learn Preprocessing
- Higher Dimensionality Boundaries

Linear Regression

- Why Squared Differences? | Footnote Section
- An Introduction to Statistical Learning | Linear Regression in Ch. 3
- Linear Regression vs PCA

Resource for Audio Machine Learning

- Voice Activity Detection in Python | MFCC Feature Extraction
- Freefield1010 Standardized Audio Data Set | Accompanying Research Paper
- Simple Minded Audio Classifier for Python