PCA - Introduction

 [(https://github.com/learn-co-curriculum/dsc-pca-introduction)](https://github.com/learn-co-curriculum/dsc-pca-introduction) [(https://github.com/learn-co-curriculum/dsc-pca-introduction/issues/new)](https://github.com/learn-co-curriculum/dsc-pca-introduction/issues/new)

Introduction

In this section, you'll learn about principal component analysis, or PCA, one of the most famous *Unsupervised Learning Techniques*. PCA is a dimensionality reduction technique. It allows you to compress a dataset into a lower dimensional space with fewer features while maintaining as much of the original information as possible.

The Curse of Dimensionality

The curse of dimensionality is a general mathematical problem relating to the exploding size of space as you continue to add additional dimensions. This can be particularly problematic when dealing with large datasets. The more features you have, the more data you have about the scenario, but the more difficult it might be to exhaustively explore combinations of these features.

PCA Use Cases

The curse of dimensionality is certainly one motivating factor for PCA. If you can't process all of the information at your disposal, then an alternative path around is necessary. Dimensionality reduction techniques such as PCA can be essential in such situations. PCA can also help improve regression and classification algorithms in many cases. In particular, algorithms are less prone to overfitting when the underlying data itself has first been compressed, reducing noise or other anomalies. Finally, PCA can also be helpful for visualizing the structure of large datasets. After all, you are limited to 2 or 3 dimensions when visualizing data. As such, reducing a dataset to 2 or 3 primary features is monumental in creating a visualization.

Summary

In this section, you'll explore PCA in depth using scikit-learn, and coding your own version from scratch using NumPy. Throughout this section, keep in mind use cases for PCA such as the curse of dimensionality.