AI Bootcamp

Principal Component Analysis (PCA)

Module 11 Day 3

Class Objectives

By the end of class, you will be able to:

1. Explain PCA and how it can be used to reduce dimensionality in data.
2. Conceptualize how PCA relates to K-means and other applications in machine learning.
3. Use PCA to reduce the number of features in an unsupervised learning setting.

Activity:

Standardizing Currency Warm-Up

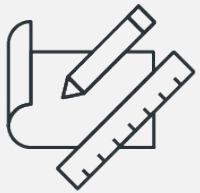
In this activity, you will use the K-means algorithm to segment global currency and interest rate data.

Suggested time:
20 minutes

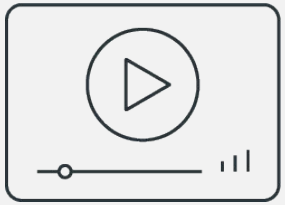
Time’s up!

Let’s review

Questions?



Introduction to **PCA**



Time for a quick **video**

Visualizing High-Dimensional Space

Introduction to PCA

Some of these datasets displayed are very large, with thousands of columns.

For some machine learning algorithms, this high dimensionality can become a problem.

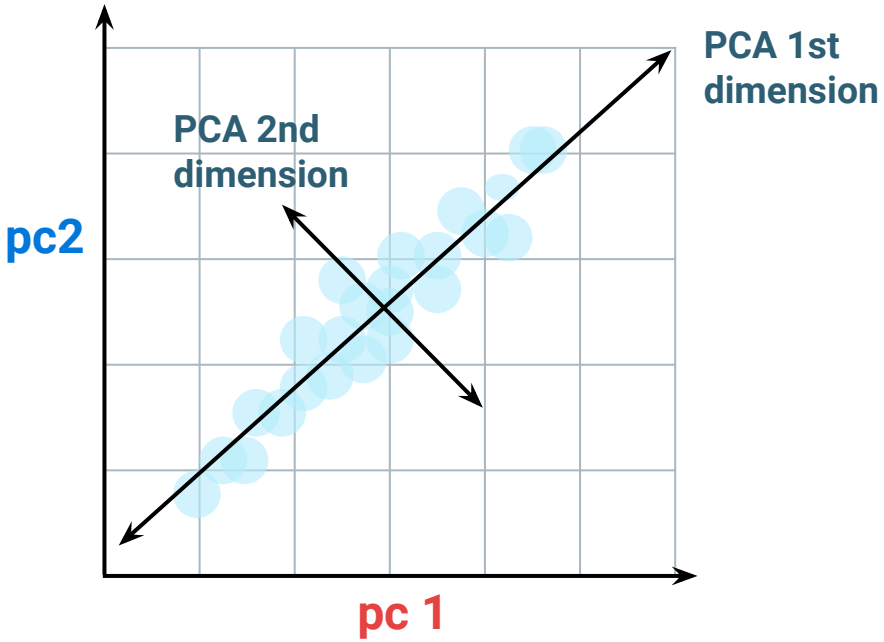
Image source

What if we could reduce the size of the dataset while preserving as much useful information as possible?

Introduction to PCA

The principal component analysis algorithm, or PCA, works because all data has **variability**; variation contains useful information.

By keeping only the most common variation across the dataset, we can reduce the size of our dataset but still maintain a variety of useful features.



Introduction to PCA

PCA is a statistical technique that we can use to streamline the machine learning process when too many factors exist in the data.

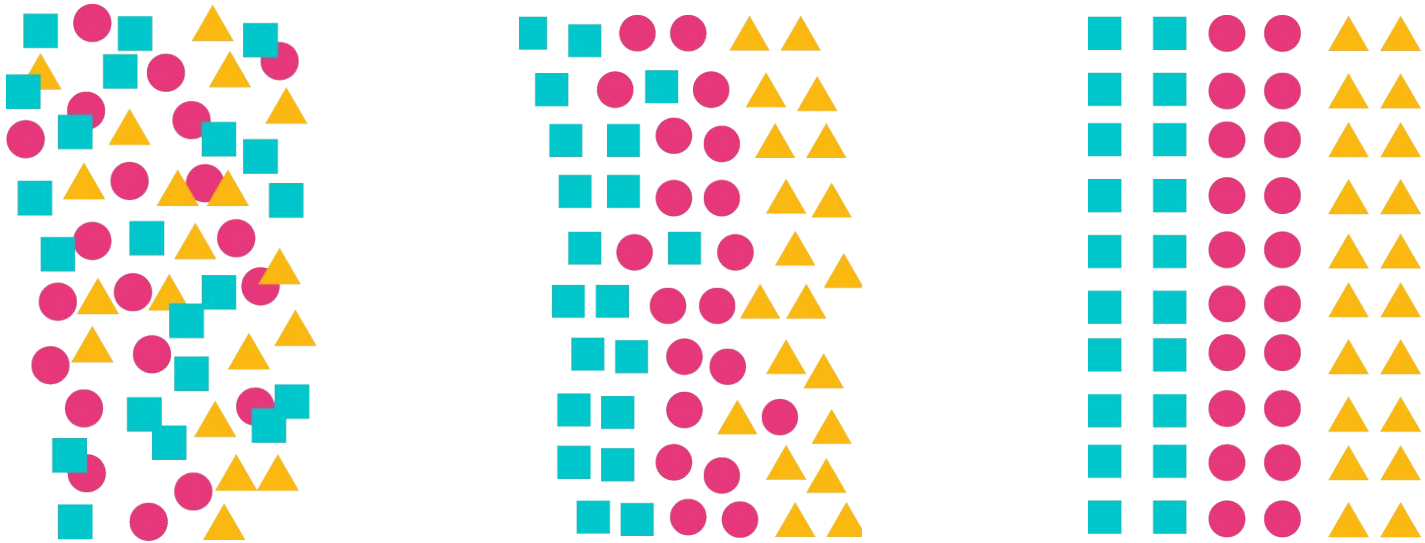
PCA reduces the number of factors by transforming a large set of features into a smaller one that contains MOST of the information of the original, larger dataset.

PCA is a dimensional reduction method that looks at all the dimensions (or data columns) in a dataset, and:

- Analyzes the weight of their contribution to the variance in the dataset.
- Reduces variables to a smaller set of dimensions that still contains as much of the information, or the maximum variance, of the original dataset as possible.

Introduction to PCA

Reducing the number of factors, or **dimensional reduction**, comes at the expense of some accuracy, but the goal is to trade a little accuracy for simplicity.



The goal of PCA is simple:
Reduce the number of factors in a dataset while preserving as much of the information from the original dataset as possible.

Instructor **Demonstration**

Introduction to PCA

Questions?

Activity:

Segmenting with PCA

In this activity, you will use your knowledge of PCA to reduce the dimensionality of a customer dataset, and determine which features have the strongest influence on the principal components.

Suggested time:
20 minutes

Time’s up!

Let’s review

Questions?

Activity:

Energize Your Stock Clustering

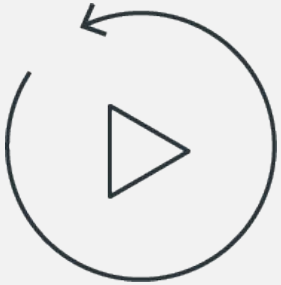
In this activity, you will use the K-means algorithm and clustering optimization techniques to cluster stocks and determine a portfolio investment strategy.

Suggested time:
20 minutes

Time’s up!

Let’s review

Questions?



Let’s **recap**

Recap

In this lesson, you learned how to:

1. Explain PCA and how it can be used to reduce dimensionality in data.
2. Conceptualize how PCA relates to K-means and other applications in machine learning.
3. Use PCA to reduce the number of features in an unsupervised learning setting.

Challenge

In this week's Challenge, you will create an unsupervised machine learning model that classifies cryptocurrencies according to their price fluctuations across various timeframes.

Questions?

The End