

Quick Review

Write a list of ways that you can do feature selection when creating a model.

Write a list of reasons why feature selection is a necessary in the machine learning project process

Introduction to Dimensionality Reduction with PCA

Aim: SWBAT explain the high-level theory of how PCA works.

Building Blocks: Feature Selection, The curse of dimensionality, Variance and Covariance

Agenda:

- Review feature selection and the curse of dimensionality
- Present problems with feature selection
- Introduce PCA
- Compare PCA to Feature Selection

The goal of dimensionality reduction

Find a low-dimensional representation of the data that retains as much information as possible.

Rules of Thumb for Feature Selection

- It's a *good thing* to have features with *high variance*, since they will be more informative and more important.
- It's a *bad thing* to have highly correlated features, or *high covariance*, since they can be deduced from one another with little loss in information, and thus keeping them together is redundant.

L

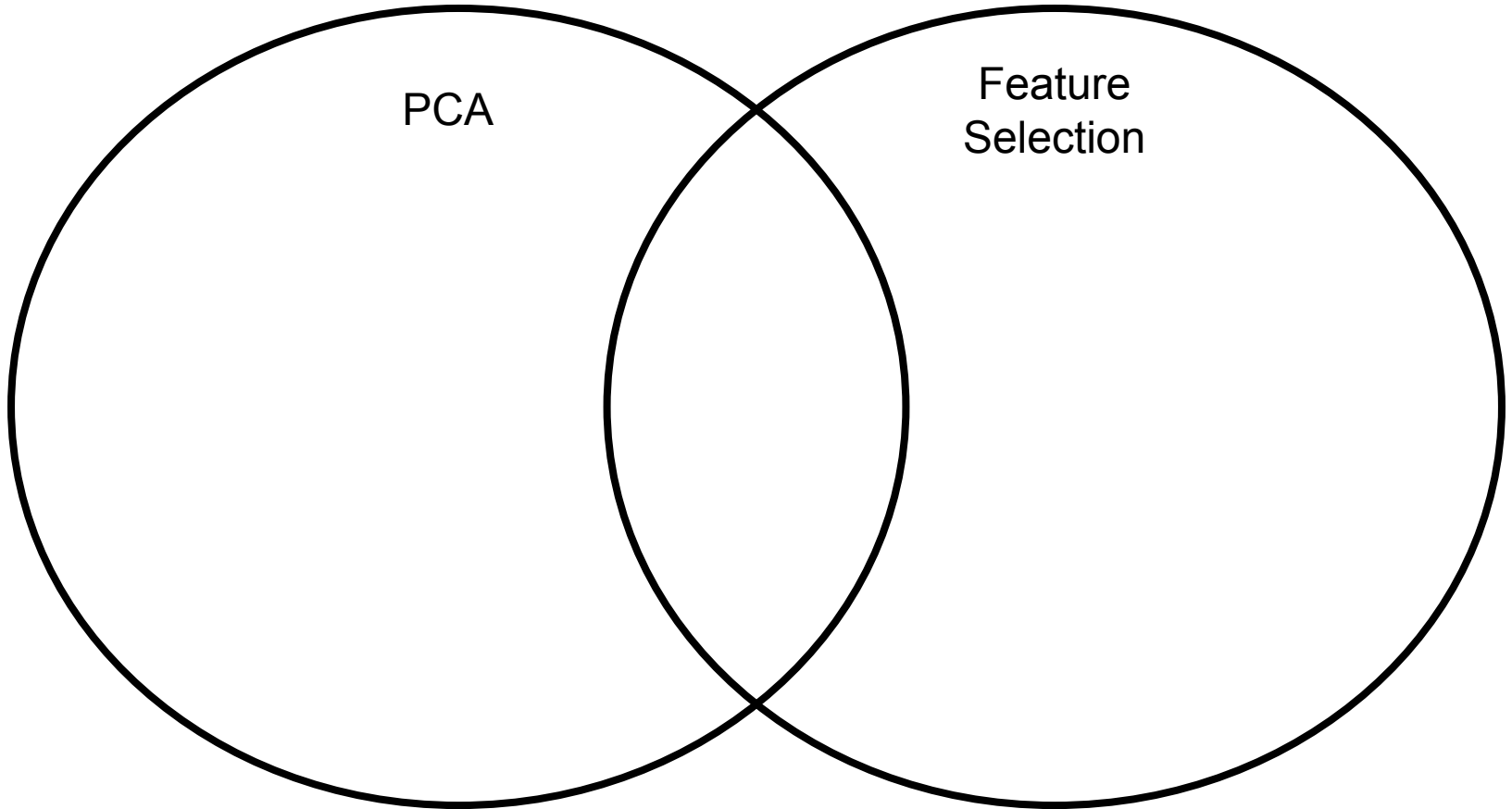
House	Floors	Square Footage	Listing Price	School Ratings	Status
1	1	1,300	\$350,000	B	Sold
2	2	1,750	\$480,000	A	On Market
3	2	2,100	\$635,000	C	Sold
4	1	1,950	\$585,000	B	On Market
5	3	2,450	\$950,000	A	Sold
6	1.5	1,800	\$480,000	B	Sold
7	2	3,000	\$775,000	C	On Market

PCA tries to find another set of axes such that the variance along that axis is as large as possible. But after rotating our axes, the axes lost their meaning—they no longer represent x_1 or x_2 . Rather, they represent a **linear combination** of both.

This rotated axis now represents a **new feature**, call it z_1 .

Likewise, the rotated x_2 -axis now represents a new feature, call it z_2 .

Comparing PCA to Feature Selection



Show me what you got

Navigate to this link and answer the question presented.

<https://goo.gl/forms/VafrcV8YhZH6dMuY2>

Next Steps

Read Parts 1-3 of “Dimensionality Reduction for Dummies”

<https://towardsdatascience.com/@abdullatif.h>

Tomorrow we will do a code implementation of PCA

Come with at least one question you have about these articles.