GENERAL ASSEMBLY

Intro to Dimensionality Reduction

Patrick Smith

LEARNING OBJECTIVES

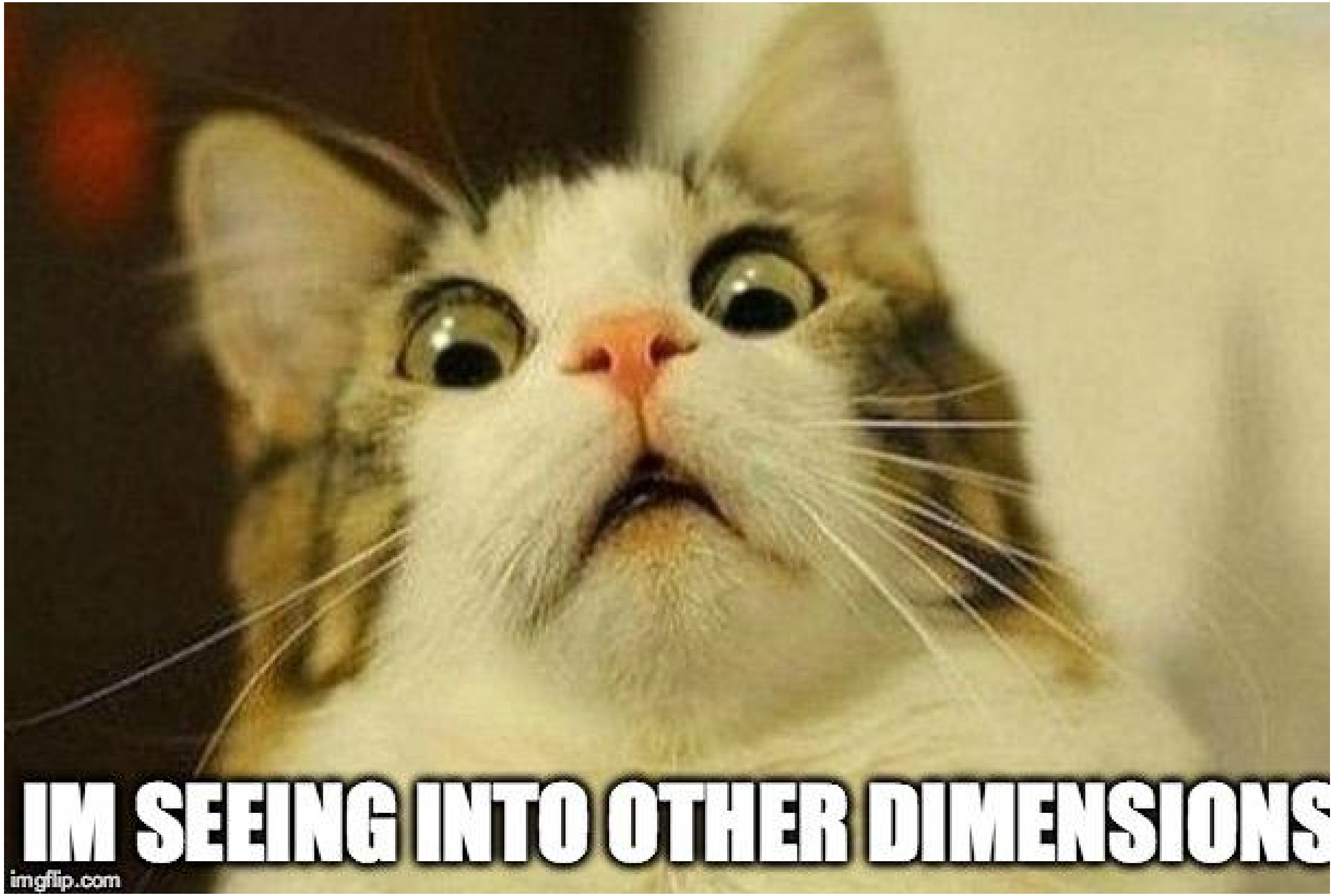
- Understand the logical workflow behind dimensionality reduction
- Understand the basic mathematical structure behind dimensionality reduction
- Calculate eigenvectors and eigenvalues for use in Principal Component Analysis

PRE-WORK

- Have a working understand of scikit learn and numpy
- Be able to create functions from scratch in python
- Have a basic understanding of linear algebra concepts such as matrices, eigenvalues, and eigenvectors

OPENING

Intro to Dimensionality Reduction



Intro to Dimensionality Reduction

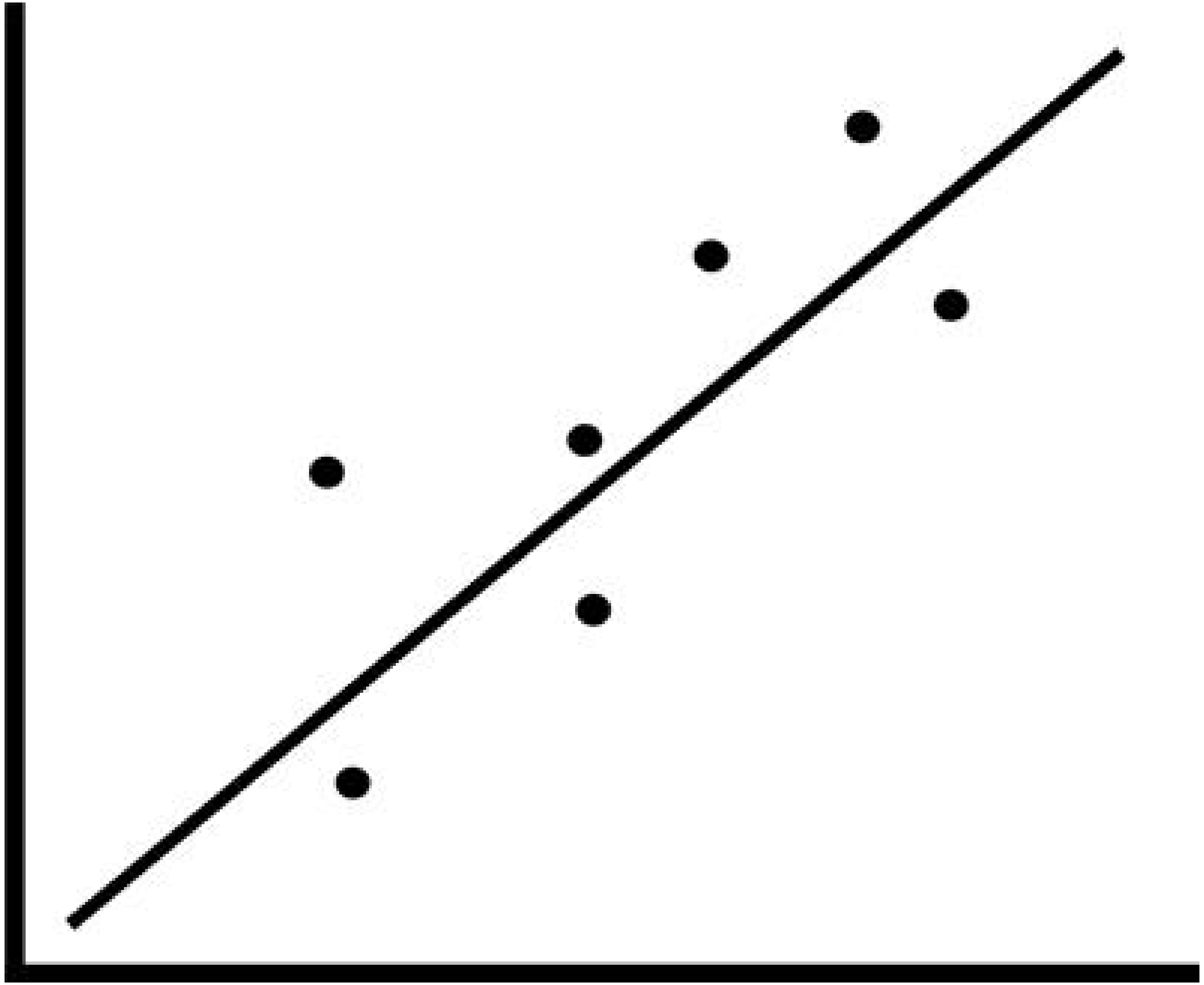
Dimensionality reduction reduces the number of random variables that you are considering for analysis until you are left with the most important variables; we want to show data in a simpler and more concise manner.

Intro to Dimensionality Reduction

Dimensionality reduction reduces the number of random variables that you are considering for analysis until you are left with the most important variables; we want to show data in a simpler and more concise manner.

Dimensionality reduction is not an end goal in itself, but a tool to form a dataset with better features for a classification or regression model.

Intro to Dimensionality Reduction



Intro to Dimensionality Reduction

Intro to Dimensionality Reduction

While dimensionality reduction can be a large piece of random forests using a feature selection approach, which you've already explored in previous lessons, we're going to focus on a form of dimensionality reduction known as ***principal component analysis*** or PCA that uses ***feature extraction.***

Intro to Dimensionality Reduction

While python has many built in applications of dimensionality reduction, such as the PCA approach that we will explore in our next lesson, today we're going to examine the the manual approach to understand the inner workings of the DR.

Dimensionality Reduction: The Longform Approach

Guided Practice: Conducting Dimensionality Reduction

Independent Practice: Dimensionality Reduction on the Iris dataset

Conclusion

Q & A

Process Review