

AI NI Academy 1

An introduction to SciKit Learn

Introductions

Jake



@JakeYoung2410

Chloe



@Chloe_Thomp

This Evening

Our First Workshop

During this workshop we will be using Google Colaboratory to take you through the basics of image classification using SciKit Learn.

Make sure that you are following along in the notebook that we have given you, relate to the slides and vice versa!

Workshop Contents

- Introduction to Machine Learning
- What is an Image?
- Understanding your data
- Data visualisation & manipulation
- Model Creation
- Model Predictions

1.

What is Machine Learning



Mat Veloso @matveloso · Nov 22

Difference between machine learning and AI:

If it is written in Python, it's probably machine learning

If it is written in PowerPoint, it's probably AI



160



6.2K

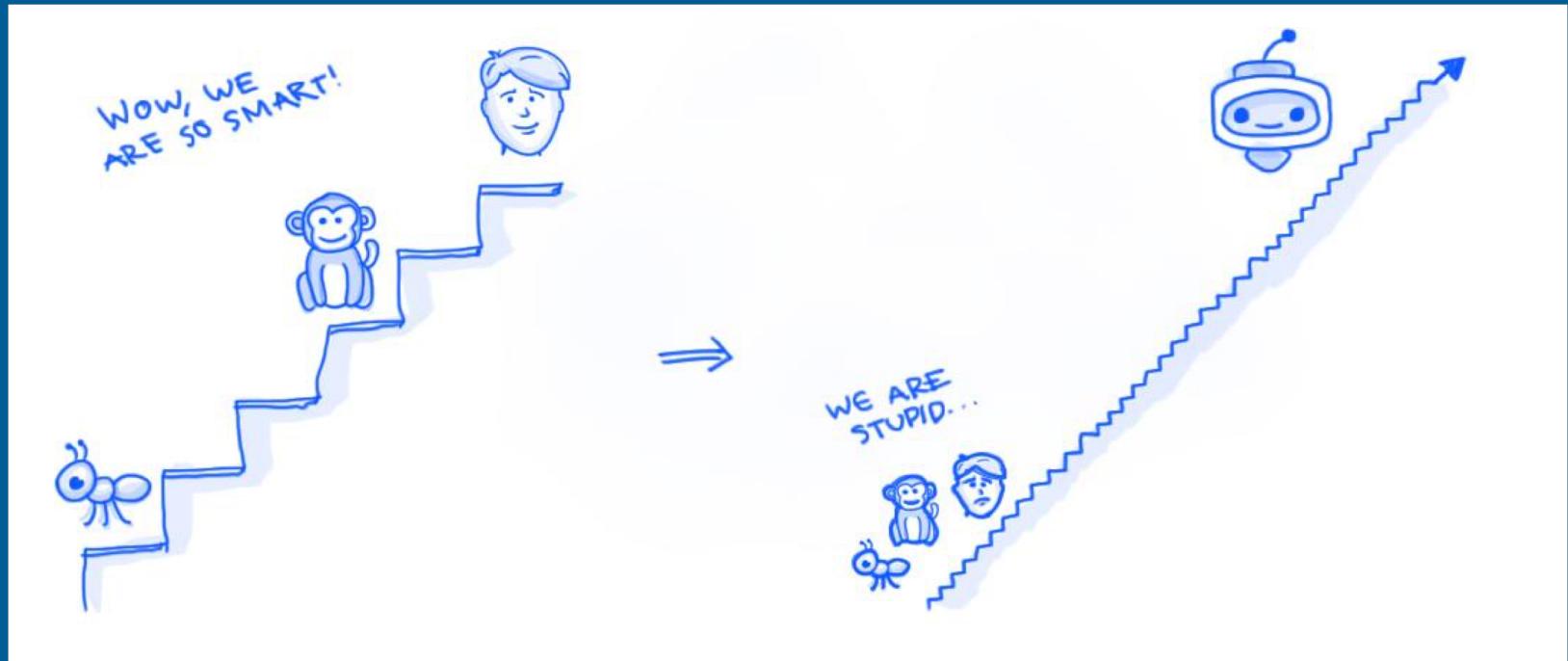


18K

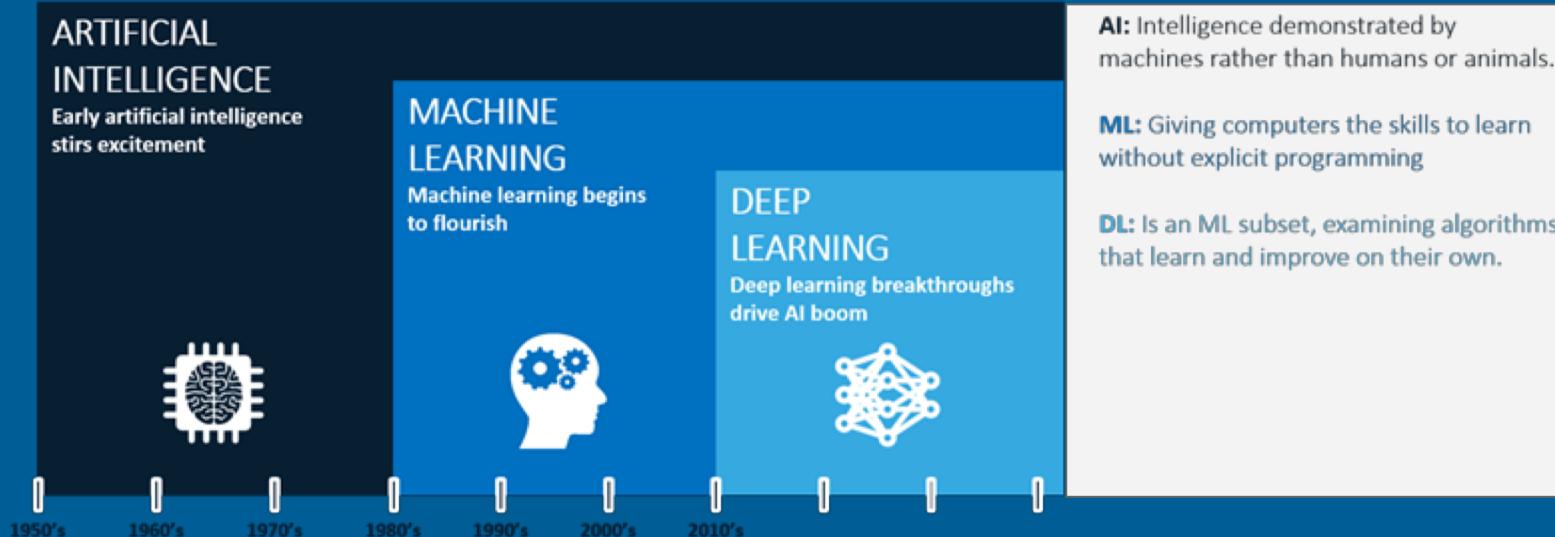


[Show this thread](#)

ANI → AGI → ASI



The Spectrum



Some Definitions

Structured Data

This type of data represents the likes of databases. Each feature has a well defined meaning.

Unstructured Data

Unstructured data covers raw audio, images and text.

Fitting

Fitting is another way of describing the training of the model.

Classes/Labels

These are the same thing. To get a prediction, we first need to train our model on correctly labelled data.

Bias

A model with high bias pays very little attention to the training data and oversimplifies the model.

Variance

A model with high variance pays a lot of attention to training data and does not generalise on the data which it hasn't seen before.

Some Definitions

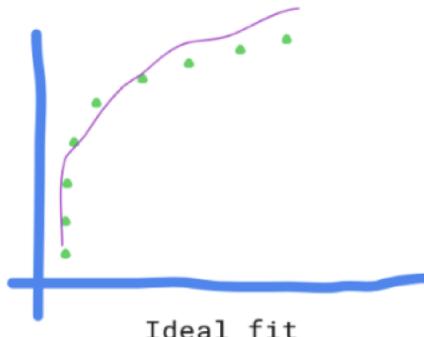
Underfitting

Due to High Bias



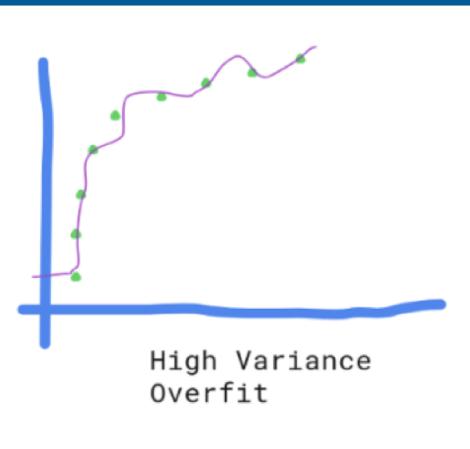
Generalising

Due to Balanced Bias and Variance



Overfitting

Due to High Variance

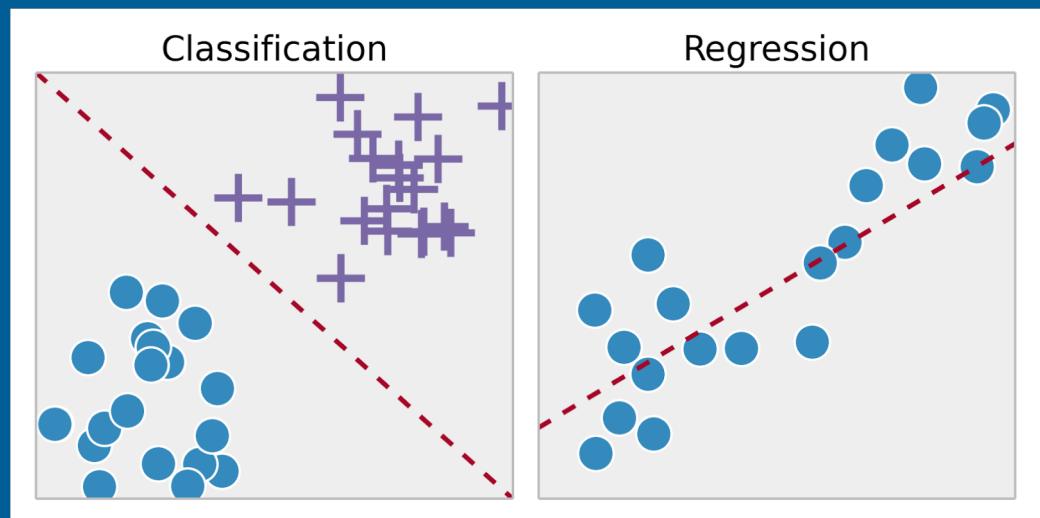


Supervised Learning

Refers to the process of building a machine learning model that is based on labelled training data.

Problems can be grouped into two groups :

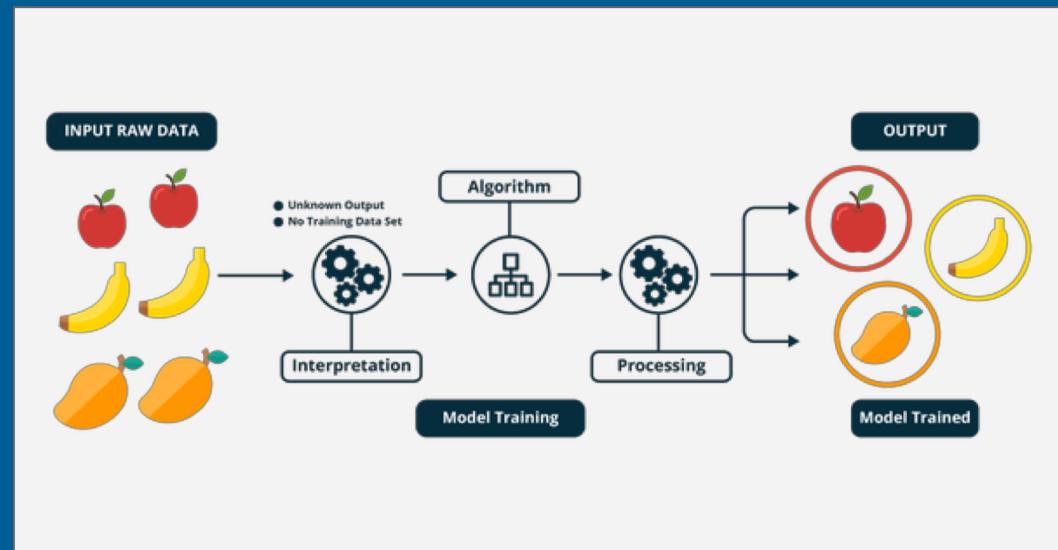
- **Classification** - The output variable is a category. E.g. "Cat" and "Dog"
- **Regression** - The output variable is a real value. E.g. "Price" or "Weight"

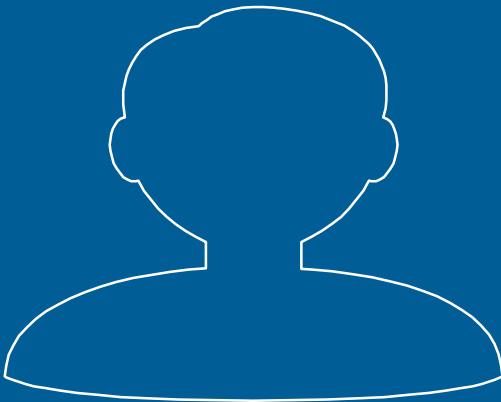


Unsupervised Learning

Unsupervised machine learning is a class of techniques to find patterns that describe the structure of “unlabeled” data (data that has not been classified or categorised).

The most common type of unsupervised learning is clustering.





Let's Learn Faces

We will be using SciKit Learn to classify faces
of famous people!

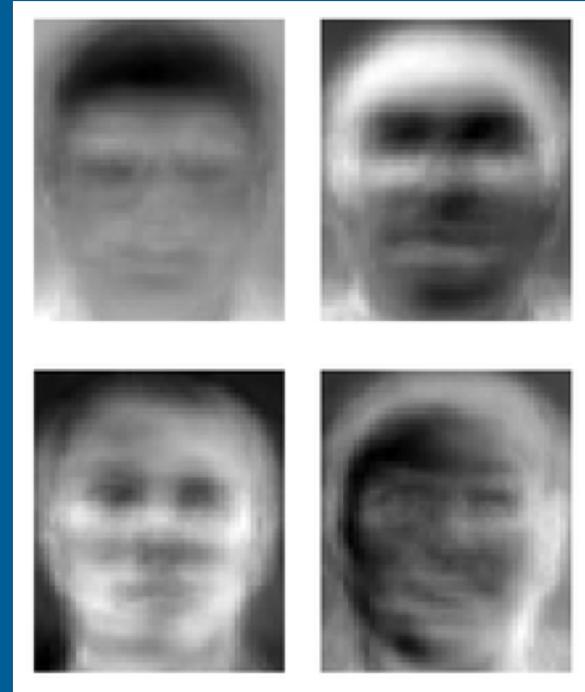
■ What is an Image?



Image Augmentation



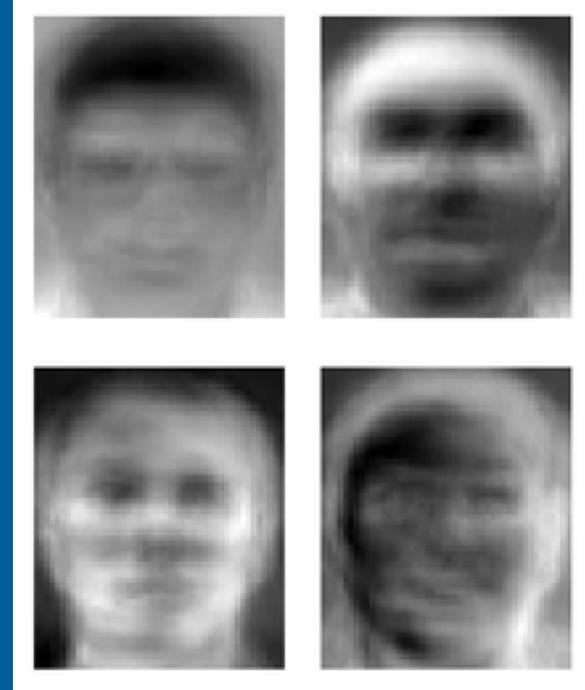
A EIGENFACE IS
WORTH A
THOUSAND
NUMBERS



■ What is an Eigenface?

An Eigenface is the fancy name we give to the vectors we use to perform facial analysis using computer vision.

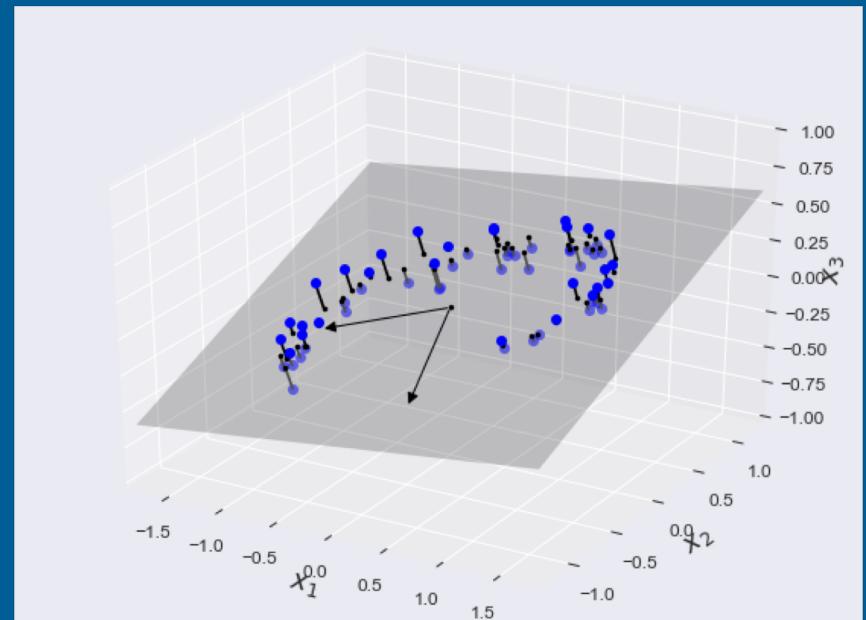
The vectors in this case are Eigenvectors which is a vector that remains unchanged when a linear transformation is applied to it.



What is PCA then?

PCA (Principal Component Analysis) is what we use to identify the hyperplane that is closest to the data points, that we then project on to so we can reduce the dimensions of our data.

We will be using this to reduce the dimensions of our data to 150.



We're only at the tip of the iceberg!
Let's Dive into Code!



THANKS!

Any questions?

You can find us at:

 @JakeYoung2410

 @Chloe_Thomp

 ai-ni-group.slack.com