Introduction to machine learning

Machine learning with Python for finance professionals

What is machine learning?

From Wikipedia:

- "Machine learning, a branch of **artificial intelligence**, is about the construction and study of systems that can **learn from data**."

What is machine learning?

From Wikipedia:

- "Machine learning, a branch of **artificial intelligence**, is about the construction and study of systems that can **learn from data**."
- "The core of machine learning deals with **representation** and **generalisation**..."

What is machine learning?

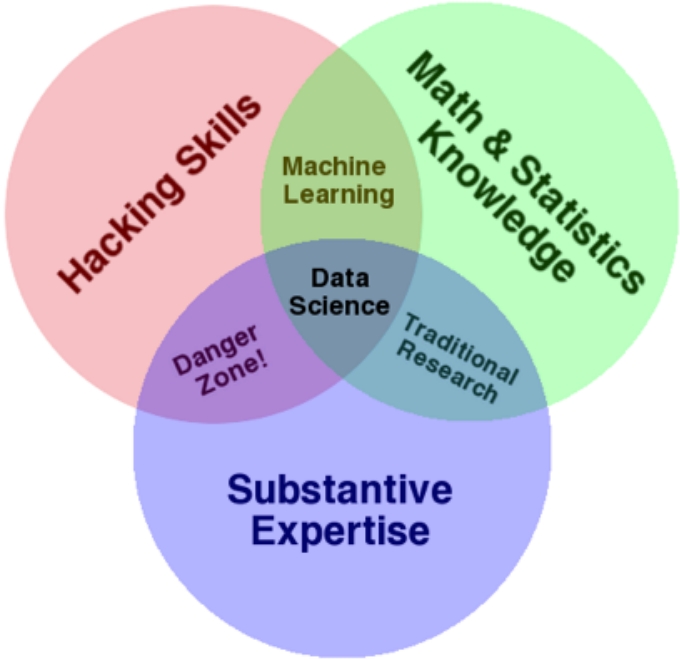
From Wikipedia:

- "Machine learning, a branch of **artificial intelligence**, is about the construction and study of systems that can **learn from data**."
- "The core of machine learning deals with **representation** and **generalisation**..."
  - **representation** – extracting structure from data
  - **generalisation** – making predictions from data

What is machine learning?

From Wikipedia:

- "Machine learning, a branch of **artificial intelligence**, is about the construction and study of systems that can **learn from data**."
- "The core of machine learning deals with **representation** and **generalisation**..."
  - **representation** – extracting structure from data
  - **generalisation** – making predictions from data



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Types of machine learning problem

supervised *making predictions*

unsupervised *extracting structure*

Types of machine learning problem

supervised *making predictions*

generalisation

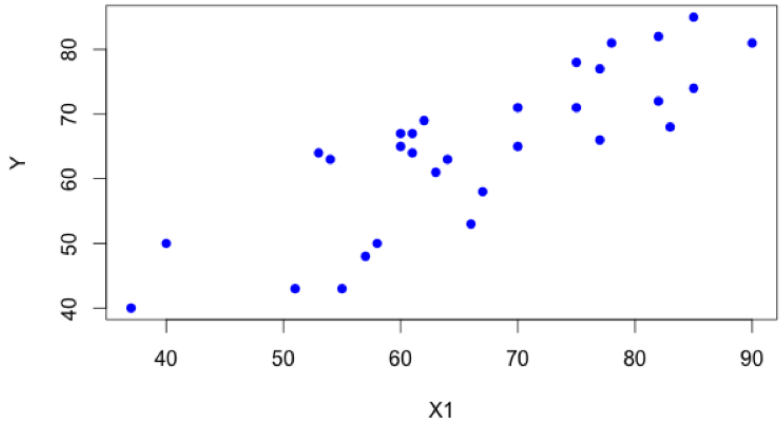
unsupervised *extracting structure*

representation

Types of machine learning problem

supervised *making predictions*

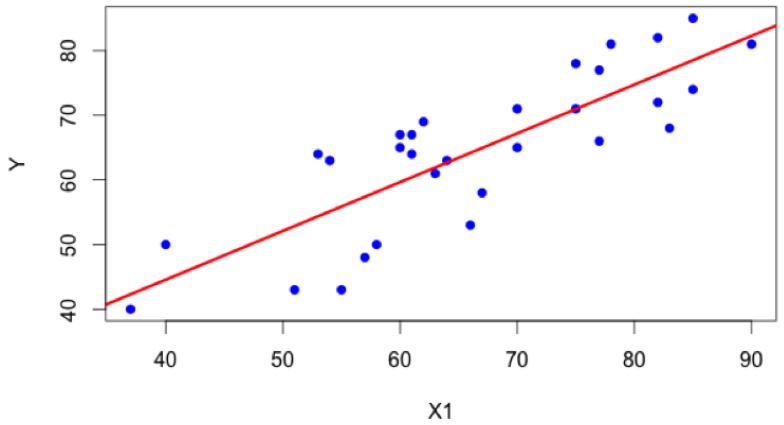
Y	X1
43	51
63	64
71	70
61	63
81	78
43	55
58	67



Types of machine learning problem

supervised *making predictions*

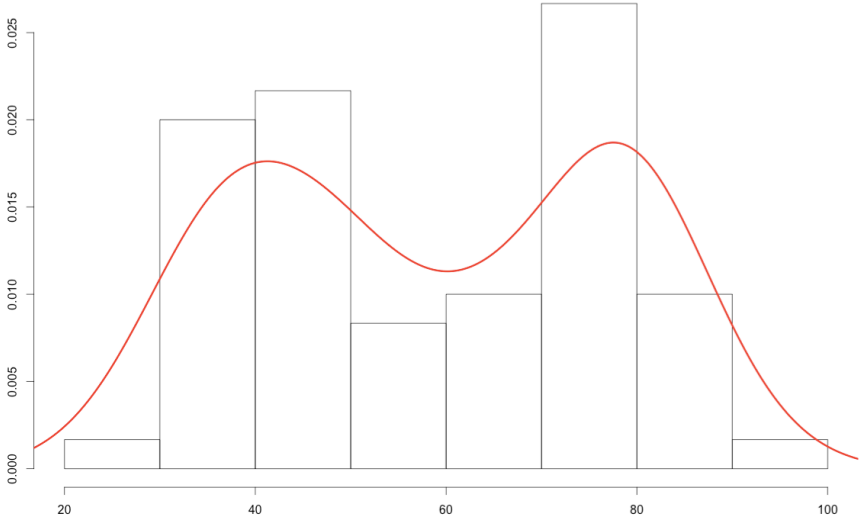
Y	X1
43	51
63	64
71	70
61	63
81	78
43	55



Types of machine learning problem

unsupervised *extracting structure*

92
73
86
84
83
49
68
66
83
80
67
74
63



Types of data

Continuous	***Categorical***
quantitative *e.g. height*	***qualitative*** *e.g. eye colour*

Types of ML problems

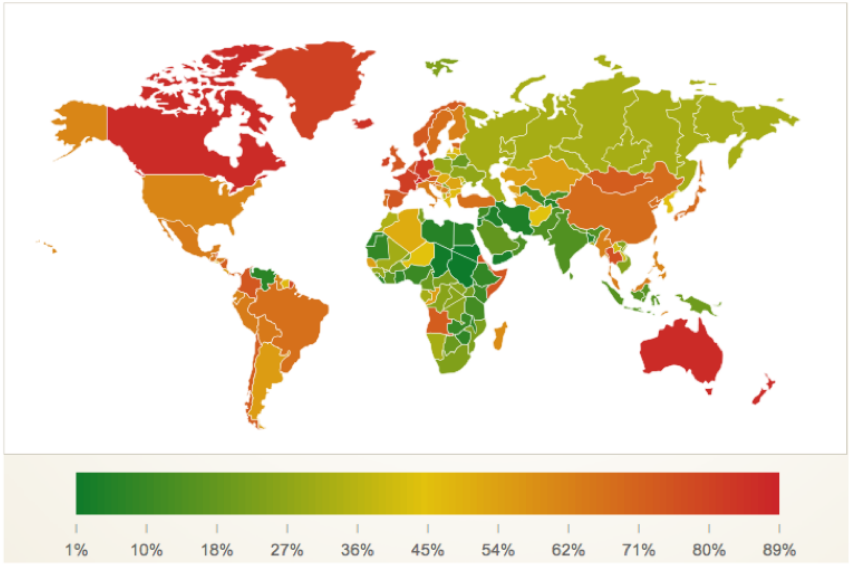
	Continuous	*Categorical*
Supervised	***Regression***	***Classification***
Unsupervised	***Dimensional reduction***	***Clustering***

Types of ML problems

	Continuous	Categorical
Supervised	Regression	Classification
Unsupervised	Dimensional reduction	Clustering

Regression example: predicting iPhone sales

- **Goal**: predict a continuous outcome variable (iPhone sales)
- **Supervised** problem + **continuous** outcome (a.k.a. dependent or target variable) implies this is a **regression** problem



GDP

population

Gini

phone penetration %

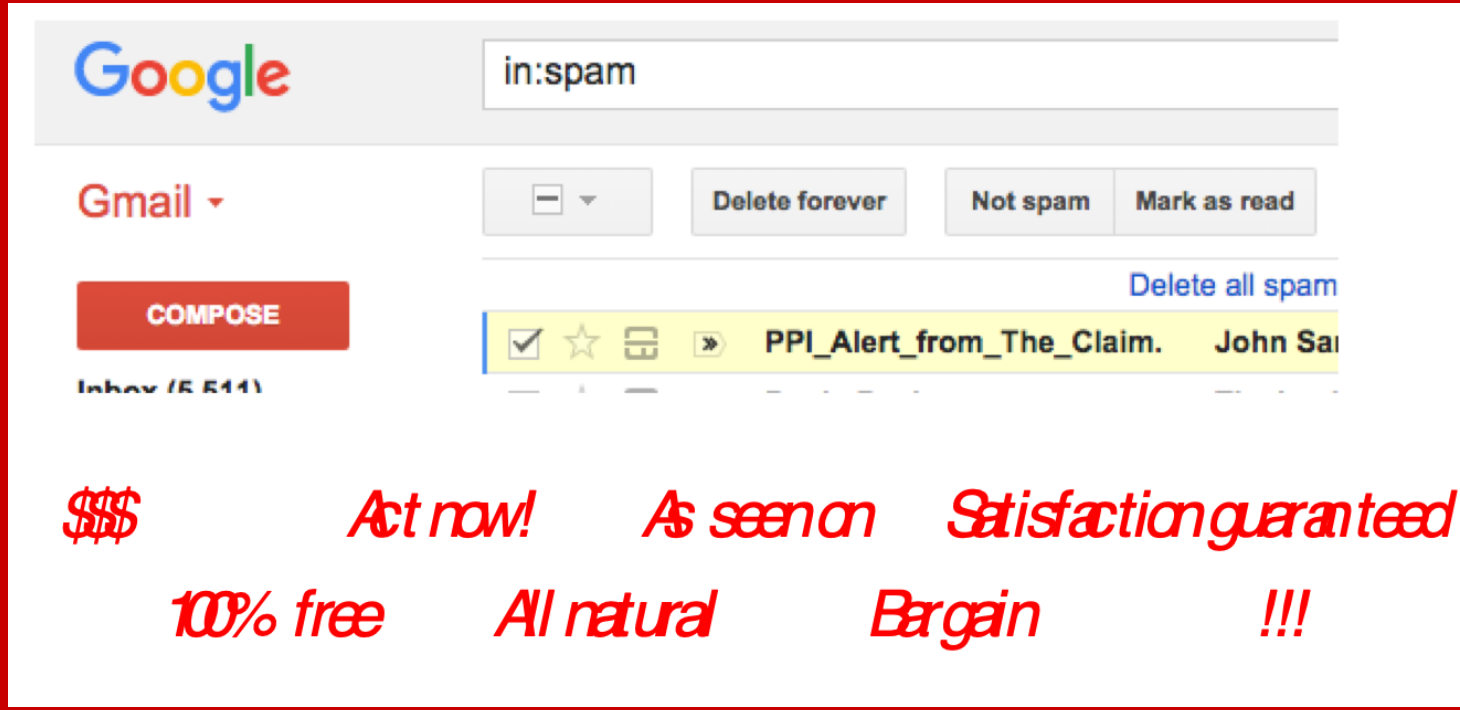
GDP growth rate

Types of ML problems

	Continuous	Categorical
Supervised	Regression	Classification
Unsupervised	Dimensional reduction	Clustering

Classification example: predicting email spam

- **Goal**: predict a **categorical** outcome variable (spam or not)
- **Supervised** problem + **categorical** outcome (a.k.a. dependent or target variable) implies this is a **classification** problem

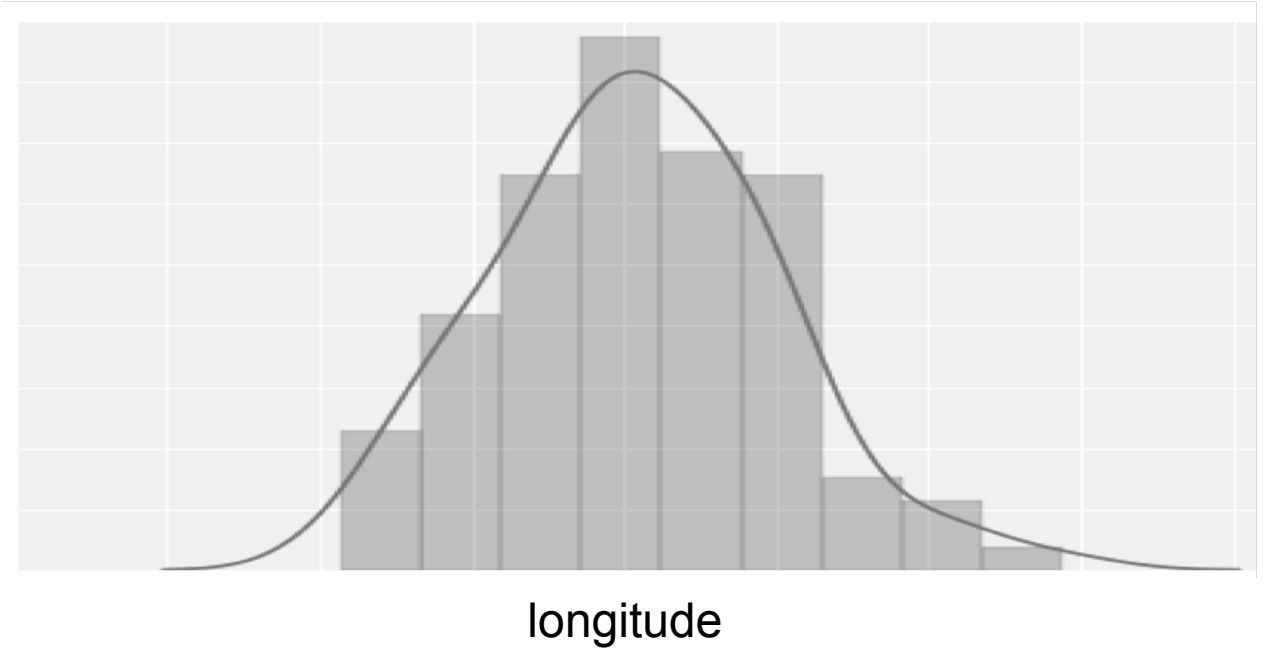


Types of ML problems

	Continuous	Categorical
Supervised	Regression	Classification
Unsupervised	Dimensional reduction	**Clustering**

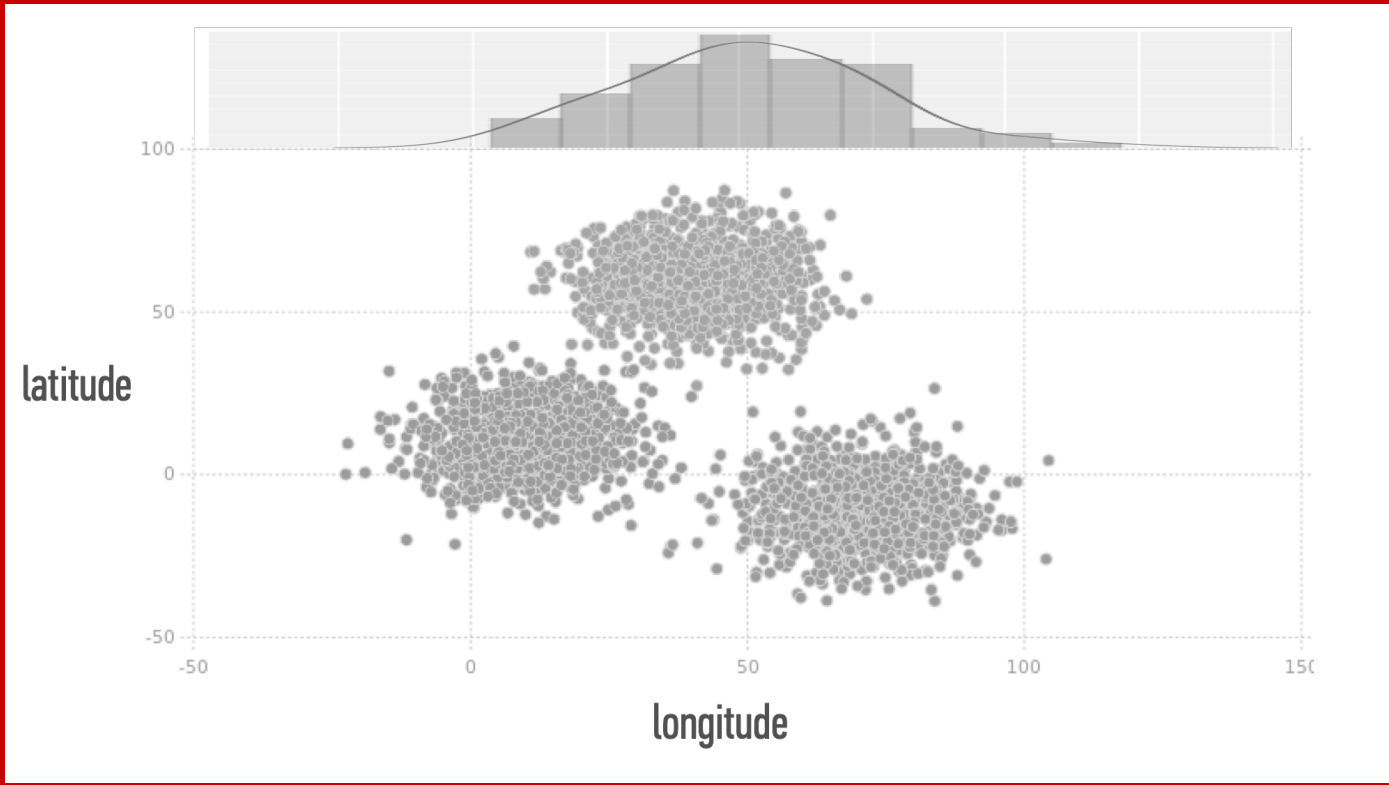
Clustering example: user locations

- Goal: look for **patterns** or **groups**
- No groupings are apparent when looking at just longitude



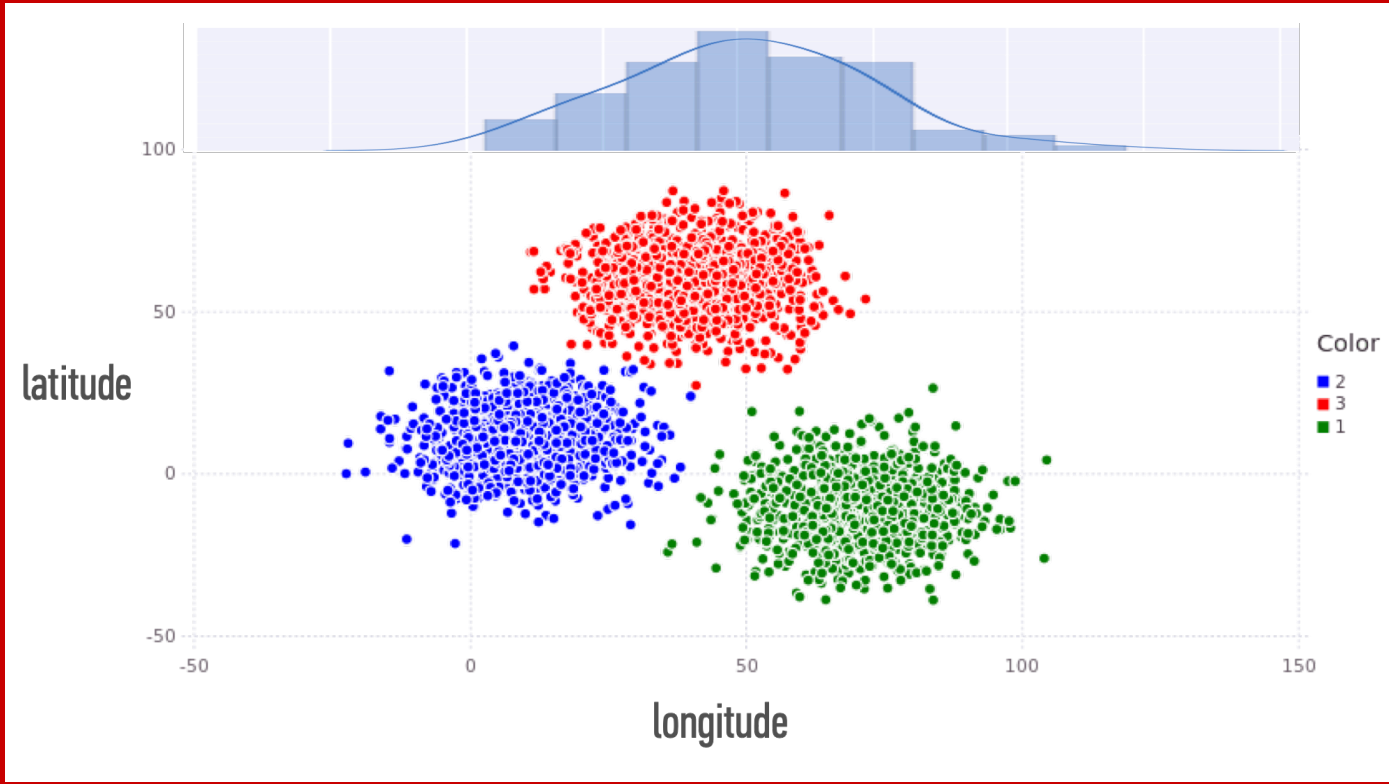
Clustering example: user locations

- Goal: look for **patterns** or **groups**
- No groupings are apparent when looking at just longitude
- Clear clusters emerge when we add latitude



Clustering example: user locations

- Goal: look for **patterns** or **groups**
- No groupings are apparent when looking at just longitude
- Clear clusters emerge when we add latitude
- Unsupervised problem + allocating each data point a **categorical group** means this is a **clustering** problem

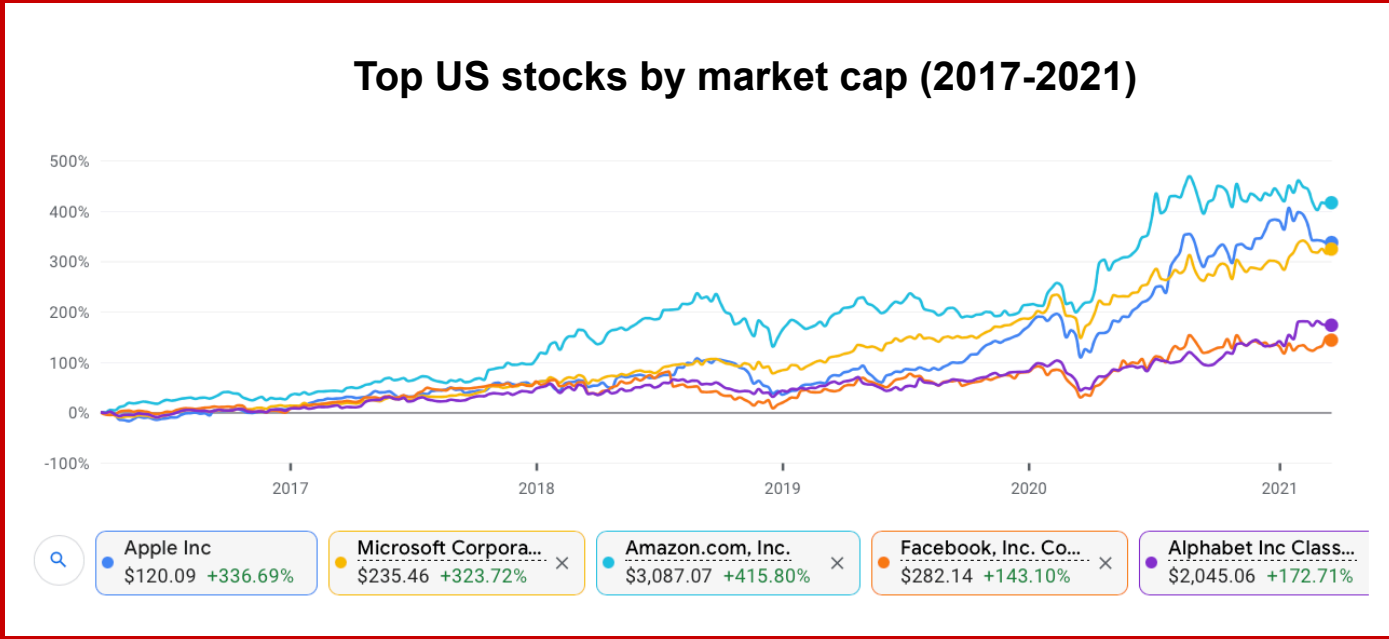


Types of ML problems

	Continuous	*Categorical*
Supervised	*Regression*	*Classification*
Unsupervised	***Dimensional reduction***	*Clustering*

Dimension reduction example: stock index

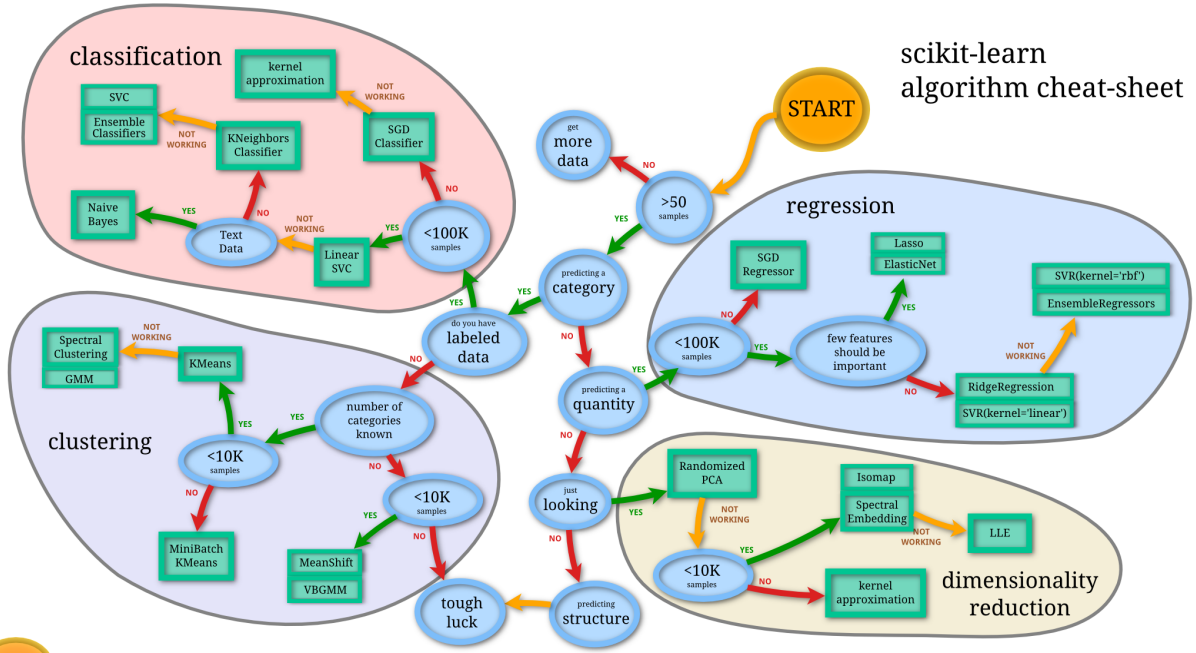
- Goal: look for **patterns** or **trends**
- Hard to discern overall trends across hundreds of stock price movements



Dimension reduction example: stock index

- Goal: look for **patterns** or **trends**
- Hard to discern overall trends across hundreds of stock price movements
- A stock index is a weighted average of its constituents' prices
- This is a **dimensional reduction** – we have reduced 500 dimensions down to one

Decision tree for algorithm selection



https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html