PLURALSIGHT

Intro to AI/ML in Azure

Welcome!

Tarek Atwan
Instructor, Pluralsight

About Me:

- Book Author
- 18+ Years Consulting Services
- 5+ Years Instructor
- 2 Startups
- World Traveler
- Gym Rat

Student Instructions

- What is your Role?
- Why did you pick this course?
- What is your experience with Azure?
- What is your AI/ML experience?
- Any Python experience?
- Any Fun fact?

Prerequisites

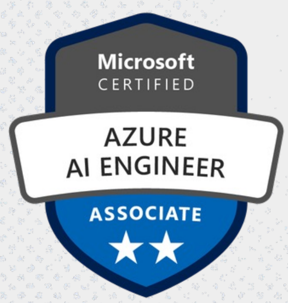
This course assumes you

- Know how to program in some other language
- Some familiarity with Python Programming
- Some familiarity with Machine Learning Concepts
- No prior experience or knowledge in Azure AI or Azure Machine Learning

Objectives

At the end of this course, you will be able to:

- Have a good understanding of Azure **Machine Learning Workspace** and **Azure AI Services**
- Be ready to continue your journey and explore taking **DP-100** and **AI-102** Certification



Microsoft Certified: Azure AI Engineer Associate

Microsoft Certified: Azure Data Scientist Associate

Azure AI and Machine Learning Services

https://azure.microsoft.com/en-us/products

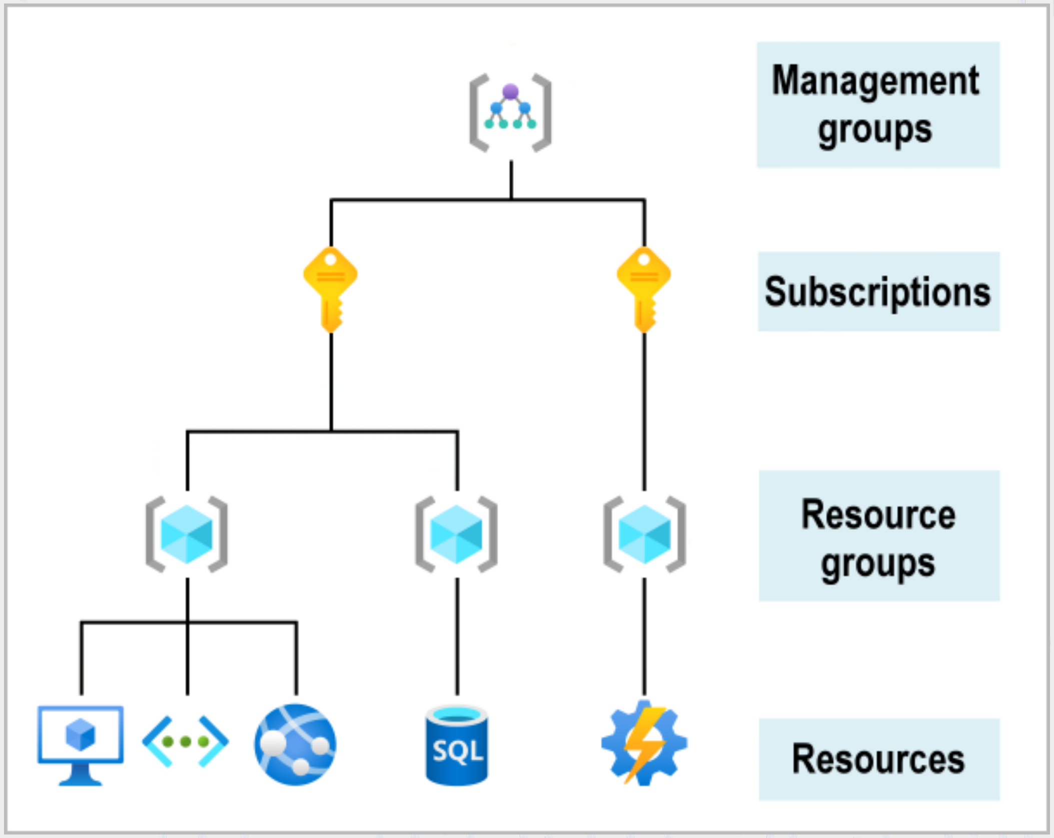
LEVELS OF SCOPE

AZURE CORE CONCEPTS

Azure Levels of Scope

Azure provides four levels of management scope:

1. Management groups
2. Subscriptions
3. Resource groups
4. Resources



Shared Responsibility Model

	Responsibility	SaaS	PaaS	IaaS	On-prem
Responsibility always retained by the customer	Information and data	Customer	Customer	Customer	Customer
	Devices (Mobile and PCs)	Customer	Customer	Customer	Customer
	Accounts and identities	Customer	Customer	Customer	Customer
Responsibility varies by type	Identity and directory infrastructure	Shared	Shared	Customer	Customer
	Applications	Microsoft	Shared	Customer	Customer
	Network controls	Microsoft	Shared	Customer	Customer
	Operating system	Microsoft	Microsoft	Customer	Customer
Responsibility transfers to cloud provider	Physical hosts	Microsoft	Microsoft	Microsoft	Customer
	Physical network	Microsoft	Microsoft	Microsoft	Customer
	Physical datacenter	Microsoft	Microsoft	Microsoft	Customer

Microsoft Customer Shared

Azure Full ML Lifecycle

Data preparation

Quickly iterate data preparation on Apache Spark clusters within Azure Machine Learning, interoperable with Microsoft Fabric.

Learn more

Feature store

Increase agility in shipping your models by making features discoverable and reusable across workspaces.

Learn more

AI infrastructure

Take advantage of purpose-built AI infrastructure uniquely designed to combine the latest GPUs and InfiniBand networking.

Learn more

Automated machine learning

Rapidly create accurate machine learning models for tasks including classification, regression, vision, and natural language processing.

Learn more

Responsible AI

Build responsible AI solutions with interpretability capabilities. Assess model fairness through disparity metrics and mitigate unfairness.

Learn more

Model catalog

Discover, fine-tune, and deploy foundation models from OpenAI, Hugging Face, and Meta using the model catalog.

Learn more

Prompt flow

Design, construct, evaluate, and deploy language model workflows with prompt flow.

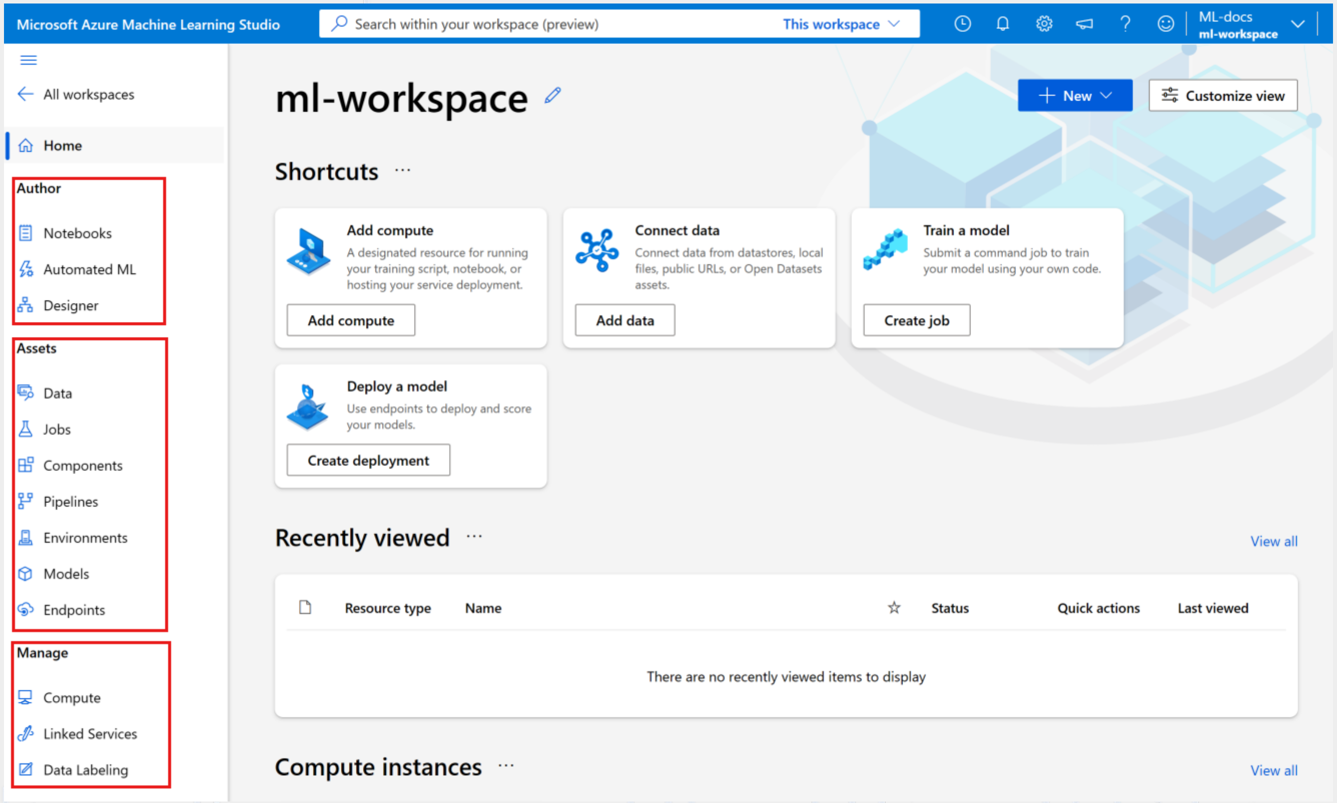
Learn more

Managed endpoints

Operationalize model deployment and scoring, log metrics, and perform safe model rollouts.

Learn more

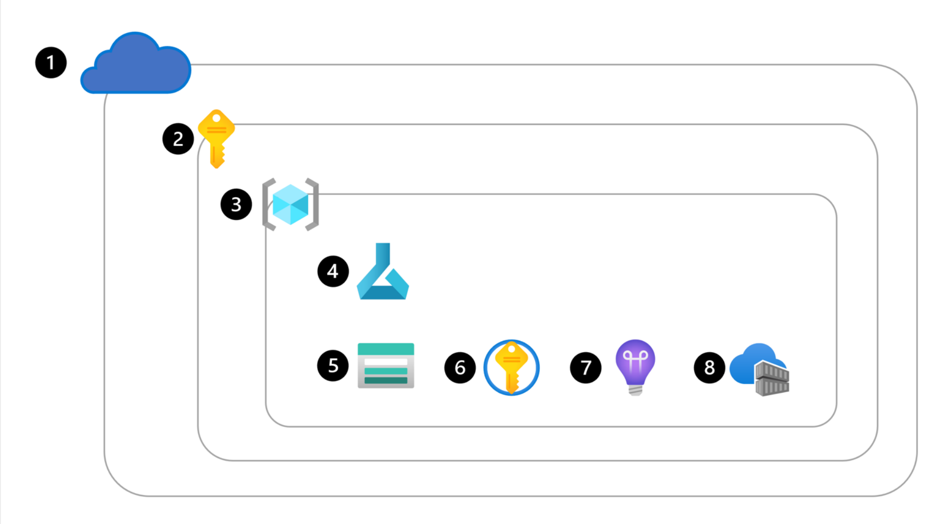
Azure ML Workspace



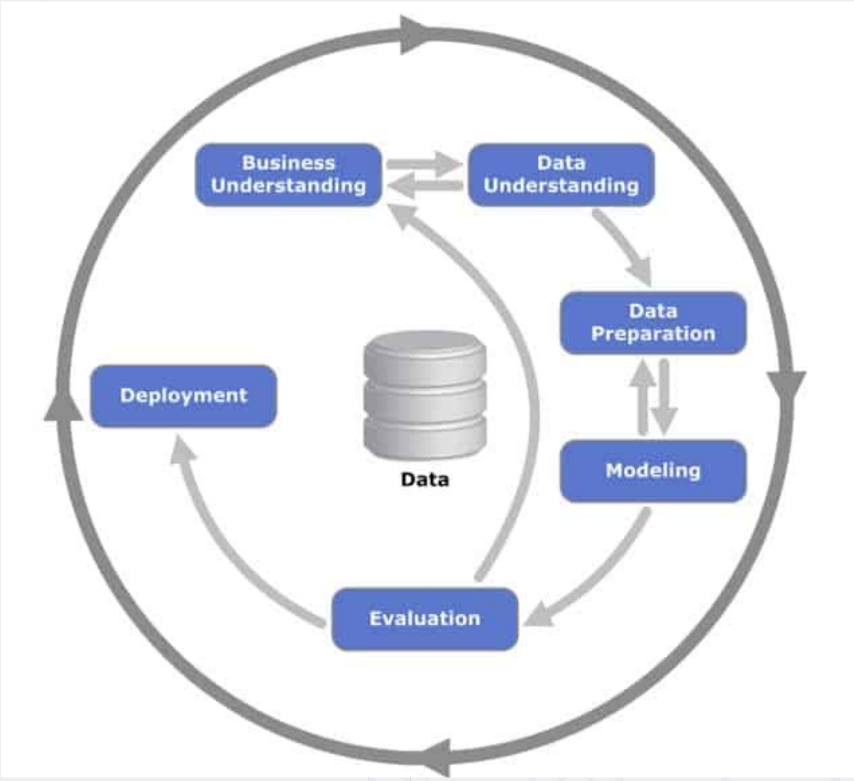
PLURALSIGHT

Azure ML Services

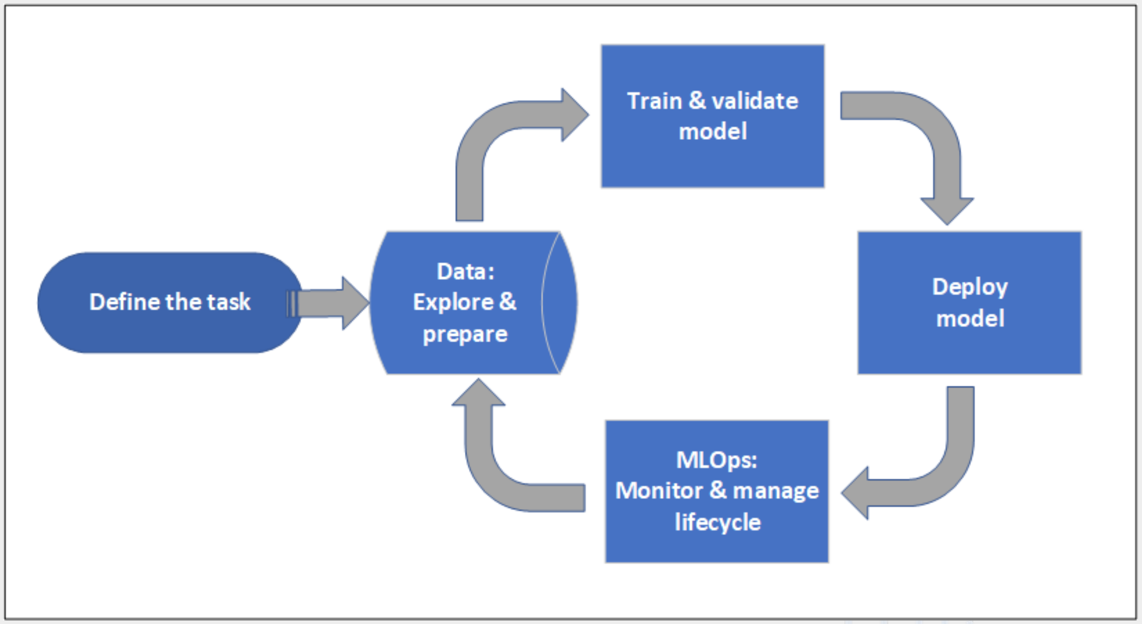
- When a **workspace** is provisioned, Azure will automatically create other Azure resources within the same resource group to support the workspace:
- **Azure Storage Account**: To store files and notebooks used in the workspace, and to store metadata of jobs and models.
- **Azure Key Vault**: To securely manage secrets such as authentication keys and credentials used by the workspace.
- **Application Insights**: To monitor predictive services in the workspace.
- **Azure Container Registry**: Created when needed to store images for Azure Machine Learning environments.



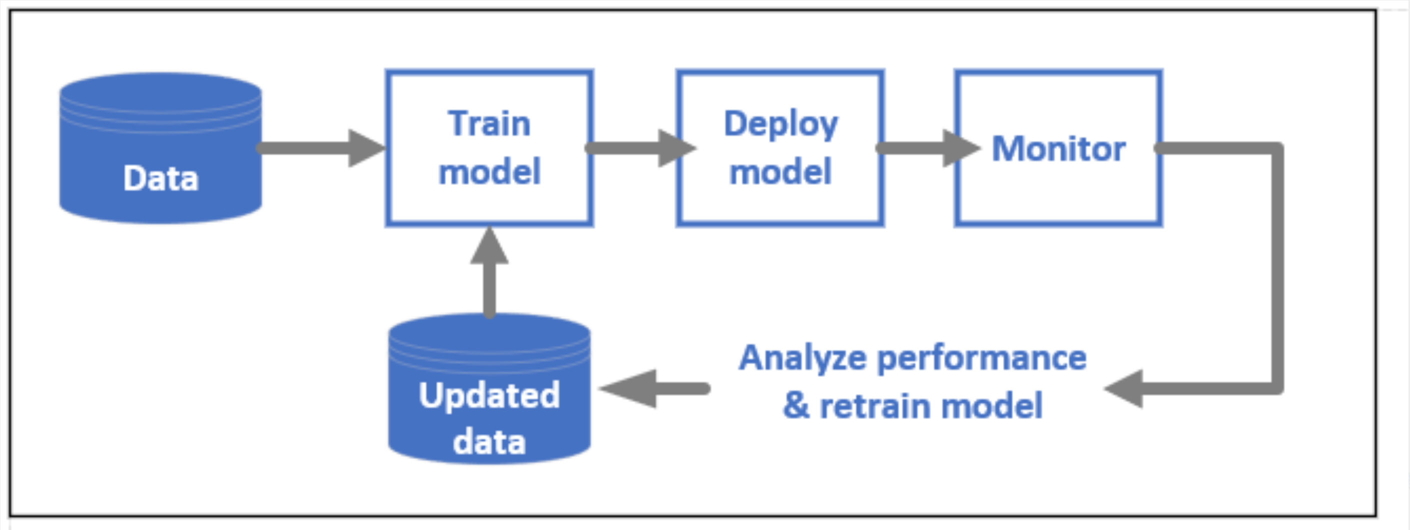
CIRSP-DM Framework



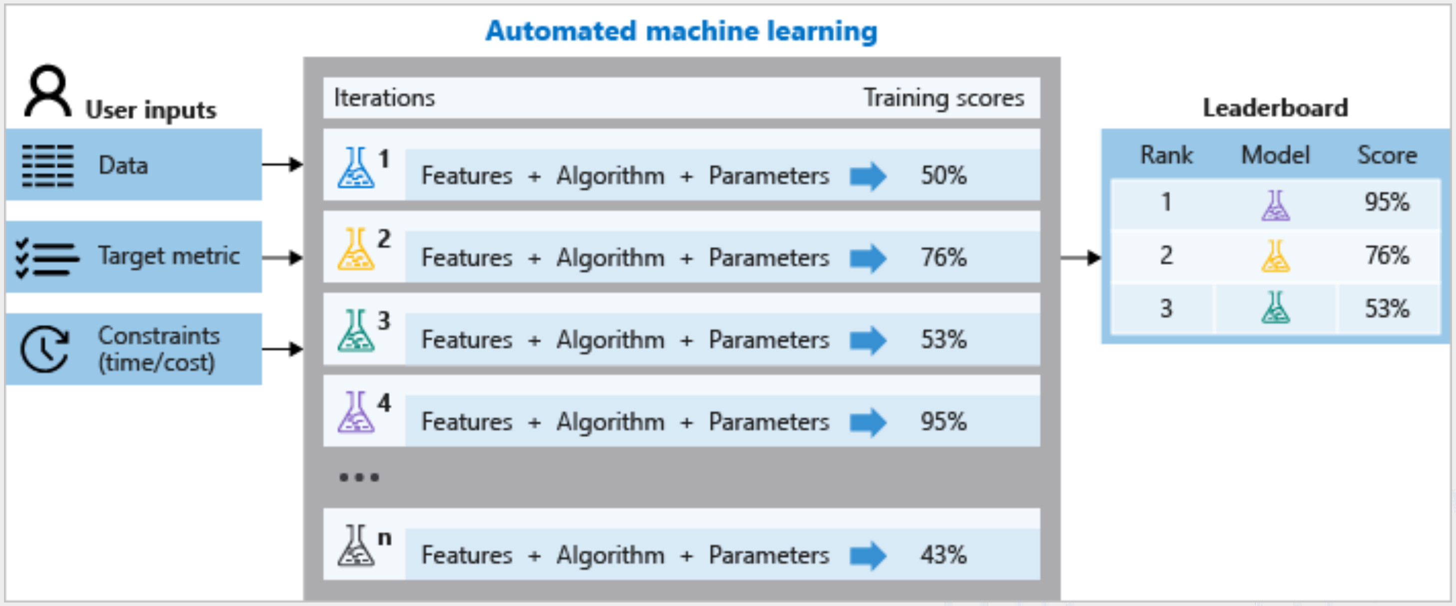
ML Project Lifecycle



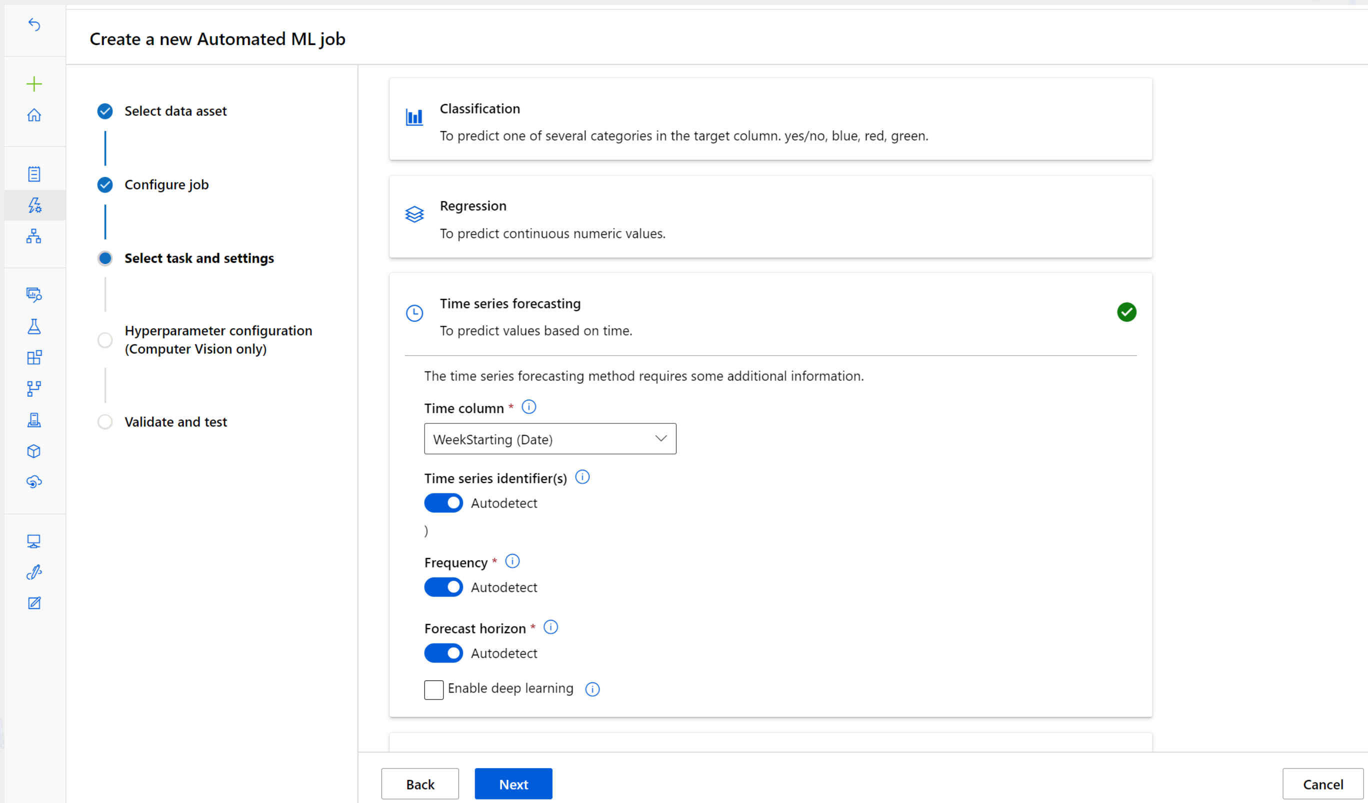
MLOps – ML Model Lifecycle



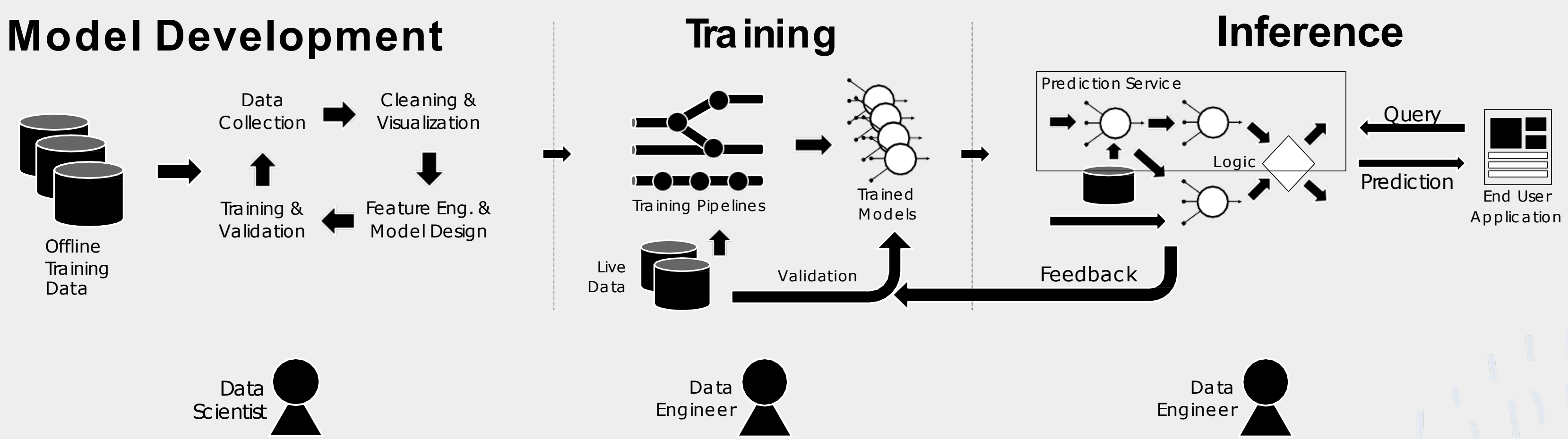
Azure Automated Machine Learning



Azure Automated Machine Learning



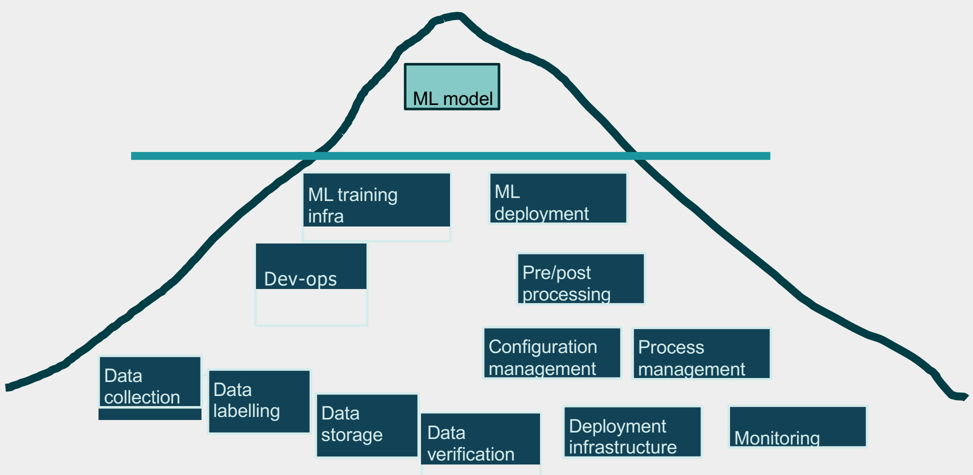
ML in Production



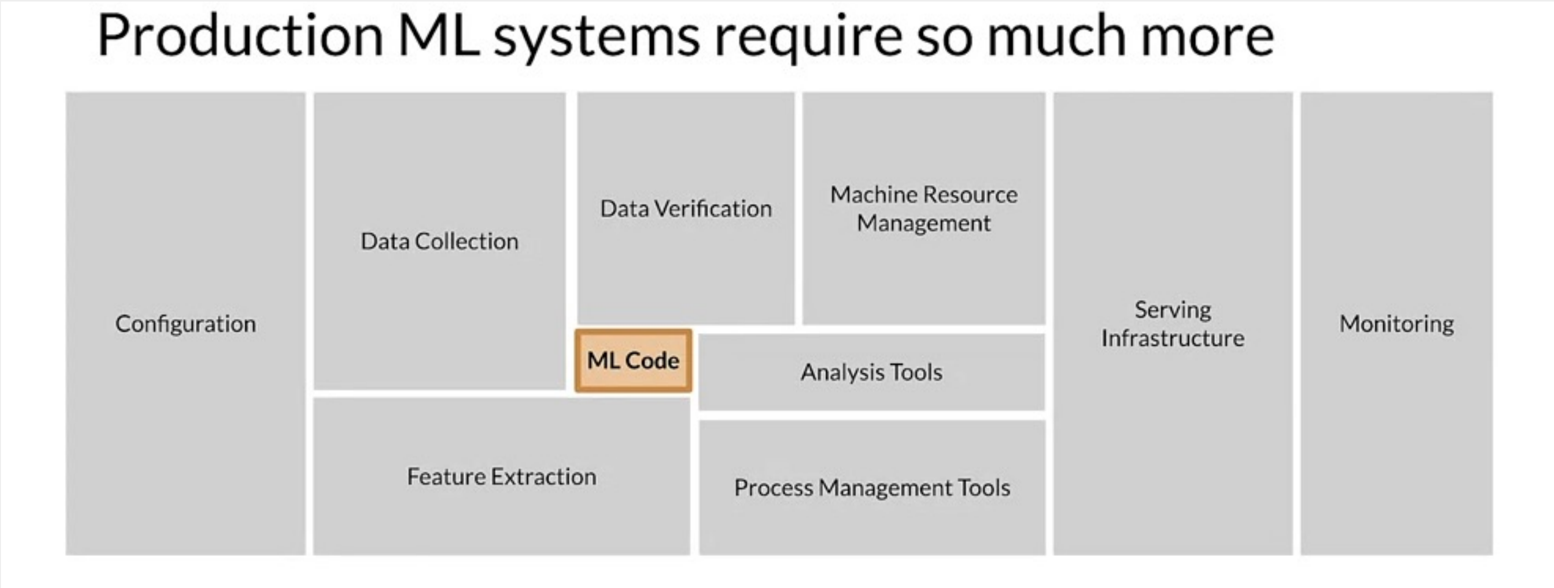
PLURALSIGHT

MLOps

Discipline that comprises a set of tools and principles to support progress through the lifecycle of a ML project



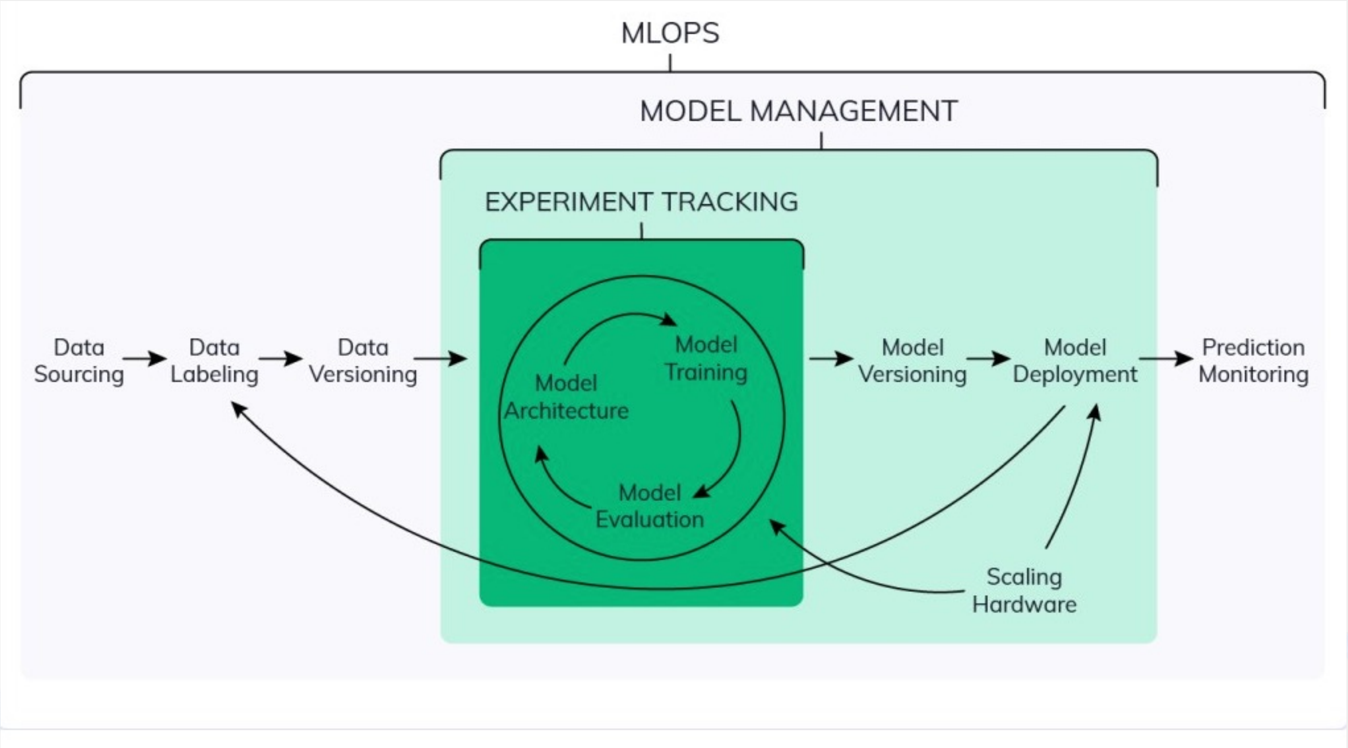
MLOps


Production ML systems require so much more
Configuration
Data Collection
Data Verification
Machine Resource Management
Serving Infrastructure
Monitoring
ML Code
Analysis Tools
Feature Extraction
Process Management Tools

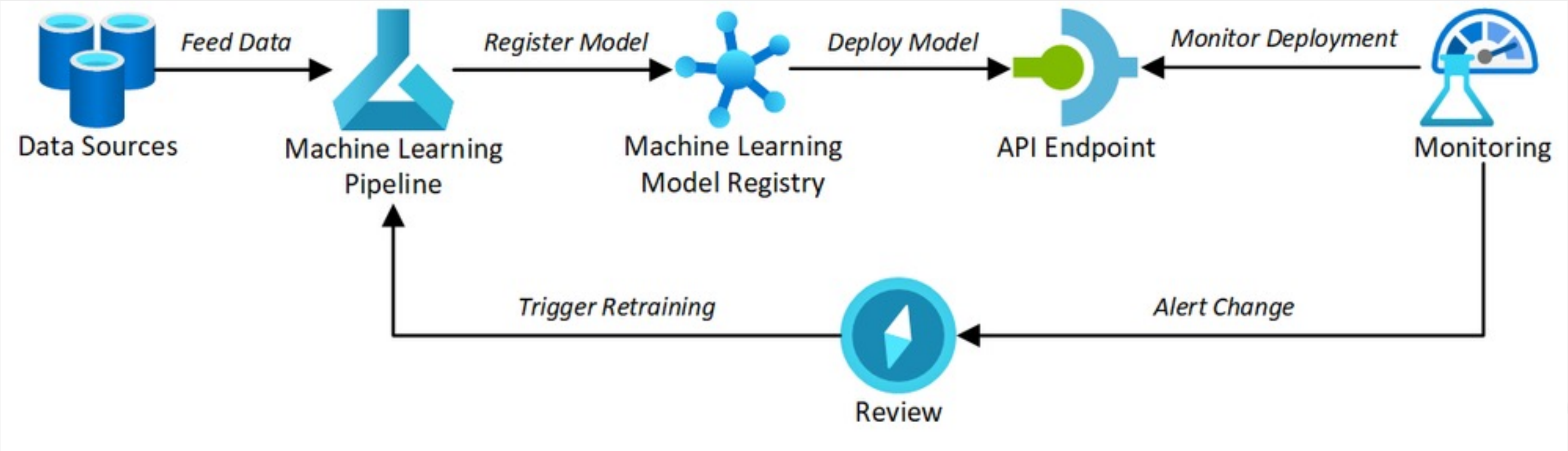
MLOps

	Academic/Research ML	Production ML
Data	Static	Dynamic - Shifting
Priority for design	Highest overall accuracy	Fast inference, good interpretability
Model training	Optimal tuning and training	Continuously assess and retrain
Fairness	Very important	Crucial
Challenge	High accuracy algorithm	Entire system

MLOps

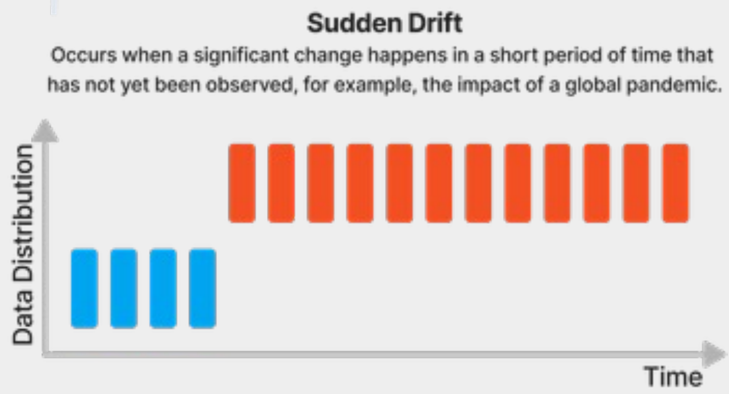

MLOPS
MODEL MANAGEMENT
EXPERIMENT TRACKING
Data Sourcing
Data Labeling
Data Versioning
Model Architecture
Model Training
Model Evaluation
Model Versioning
Model Deployment
Prediction Monitoring
Scaling Hardware

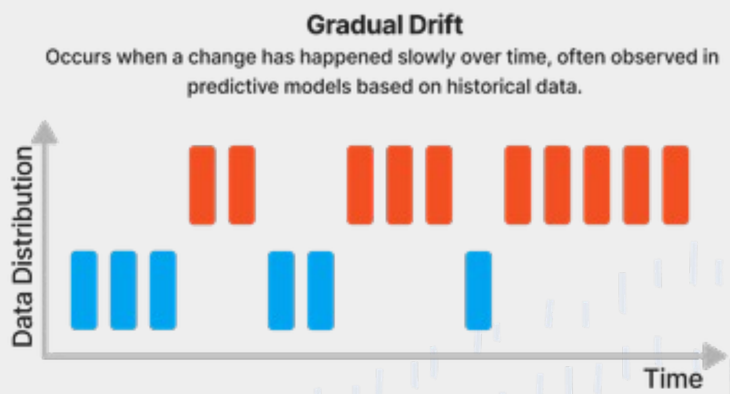
Model Drift



Data Sources
Feed Data
Machine Learning Pipeline
Register Model
Machine Learning Model Registry
Deploy Model
API Endpoint
Monitor Deployment
Monitoring
Alert Change
Review
Trigger Retraining

Model Drift

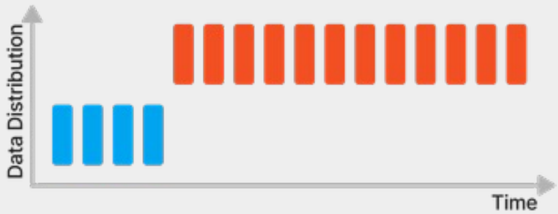

Sudden Drift
Occurs when a significant change happens in a short period of time that has not yet been observed, for example, the impact of a global pandemic.
Data Distribution
Time


Gradual Drift
Occurs when a change has happened slowly over time, often observed in predictive models based on historical data.
Data Distribution
Time

Model Drift

Sudden Drift

Occurs when a significant change happens in a short period of time that has not yet been observed, for example, the impact of a global pandemic.

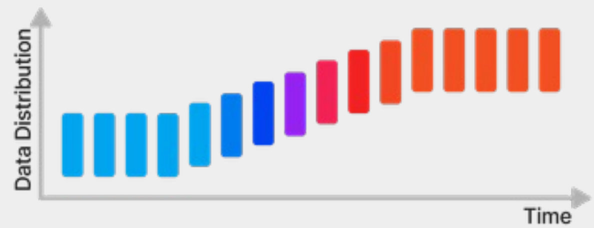


Gradual Drift

Occurs when a change has happened slowly over time, often observed in predictive models based on historical data.

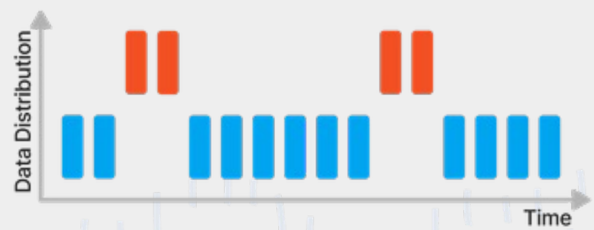
Incremental Drift

Occurs when the change is not continuous, such as predicting sales of a specific product that changes in the future.



Reoccurring Concepts

Occurs in repeating patterns, for example, predicting seasonal sales of products such as winter coats.



What is Artificial Intelligence

Artificial Intelligence

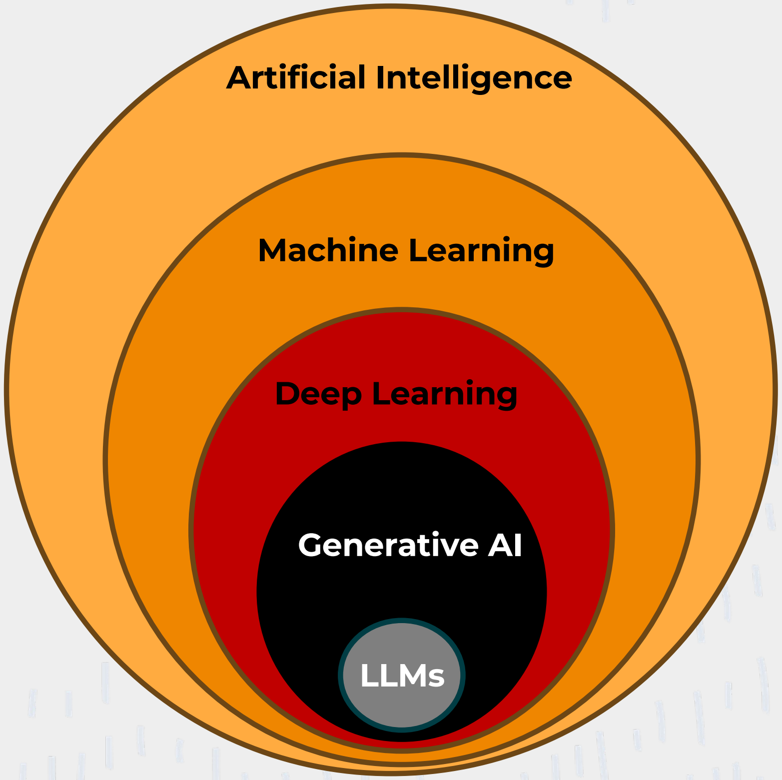
AI refers to the broad concept of machines or computers performing tasks that typically require human intelligence. This includes reasoning, learning, problem-solving, perception, language understanding, etc.

ML is a subset of AI focused on the idea that machines can learn from data, identify patterns, and make decisions with minimal human intervention

DL is a subset of ML that uses neural networks with many layers (deep networks) to model complex patterns in data.

Generative AI refers to a class of AI, often realized through DL, that focuses on generating new content or data that is similar to but distinct from the training data.

LLMs are a type of deep learning model designed to understand, generate, and interact with human language at a large scale. They are trained on vast amounts of text data.



Machine Learning Categories

Machine Learning

Unsupervised Learning

Dimensionality Reduction

- Meaningful Compression
- Big Data Visualization
- Structure Discovery
- Feature Elicitation

Clustering

- Recommender Systems
- Targeted Marketing
- Customer Segmentation

Supervised Learning

Classification

- Image Classification
- Identity Fraud Detection
- Customer Retention
- Diagnostics

Regression

- Population Growth Prediction
- Advertising Popularity Prediction
- Weather Forecasting
- Market Forecasting
- Estimating Life Expectancy

Reinforcement Learning

- Real-time decisions
- Robot Navigation
- Game AI
- Skill Acquisition

Supervised learning comprises algorithms for which the potential outcomes are knowable in advance (i.e., category or numeric range) and which we can use to correct the model's predictions.

Unsupervised Learning: Algorithms for which the potential outcomes are unlabeled. Inferences are made directly from the data without feedback from known outcomes or labels.

Machine Learning (Supervised)

Using data such as credit score, credit history, income, etc., we are trying to predict whether an individual is a credit risk or not.

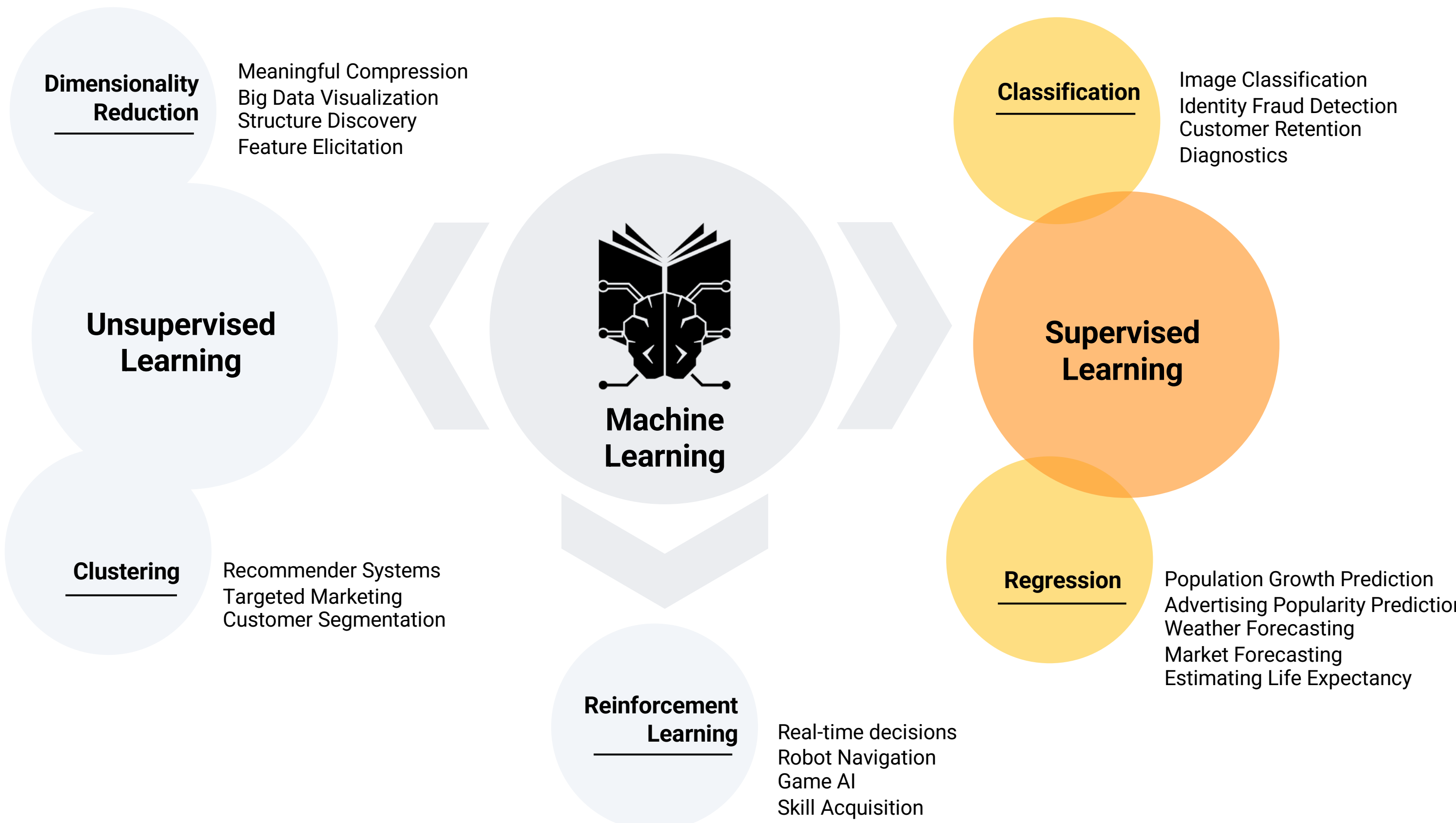
Known Category:
"Credit Risk" vs. "Not Credit Risk"

Machine Learning (Supervised)

Using features such as number of bedrooms, square feet, etc., we are trying to predict the market value of a house.

Numeric Range:
$50,000—$500,000

Supervised Learning Subcategories

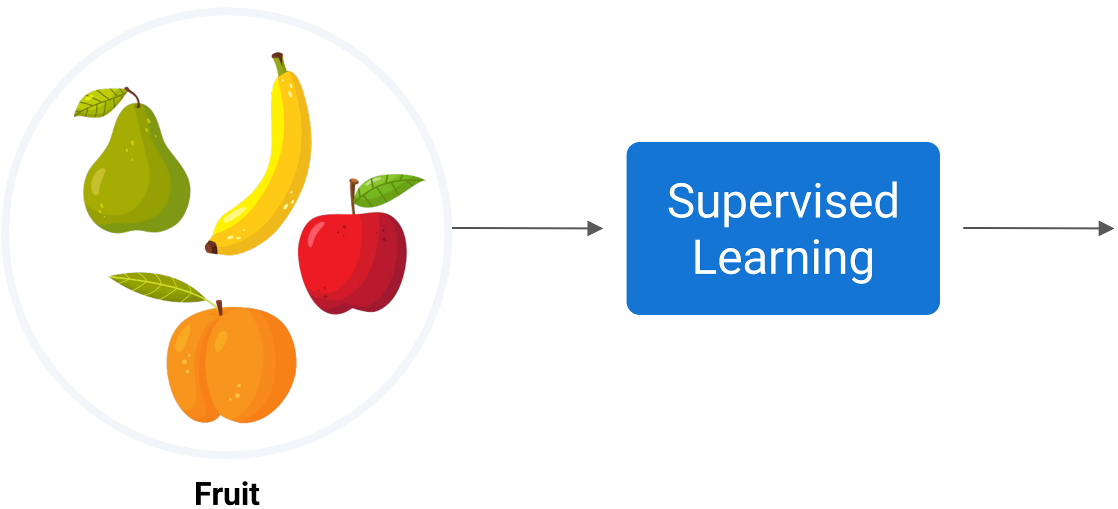


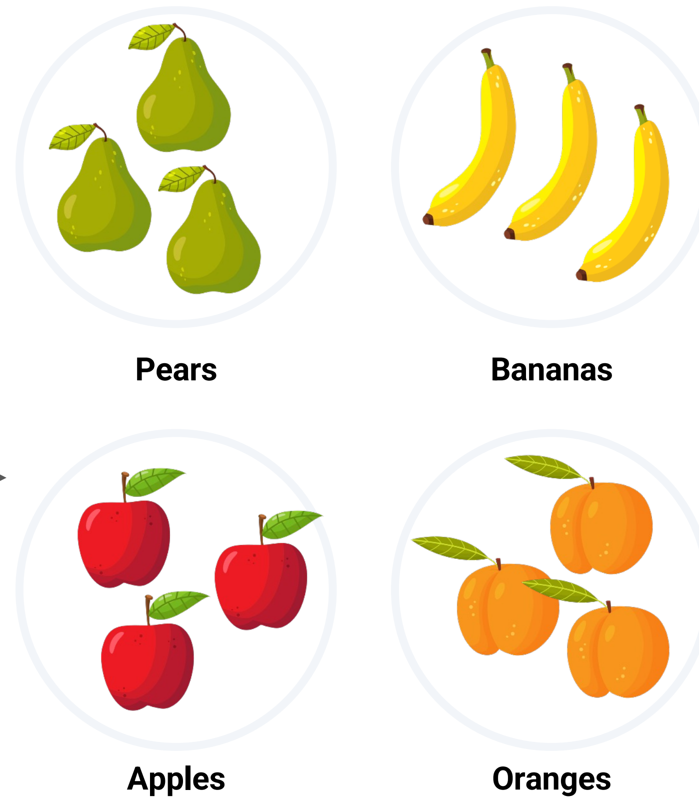
Introduction to Supervised Learning & Classification

Introduction to Supervised Learning

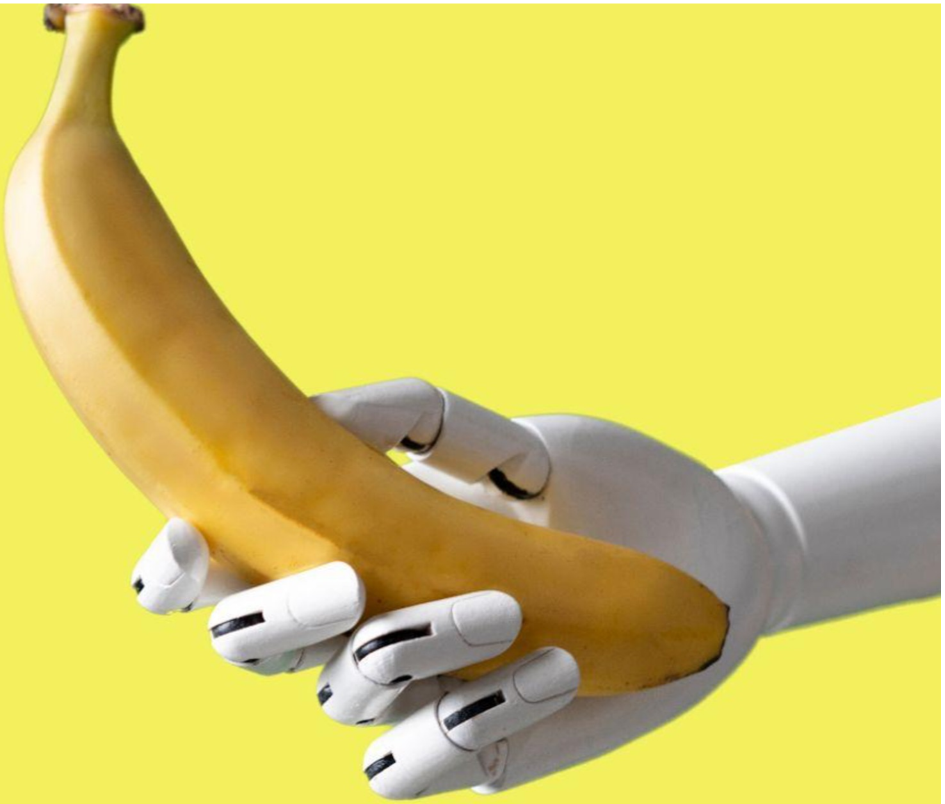
In supervised learning, we take a set of known answers called **labels** and fit a model with a set of **features** (inputs) that corresponds to the labels.

These models are called **supervised learners**.



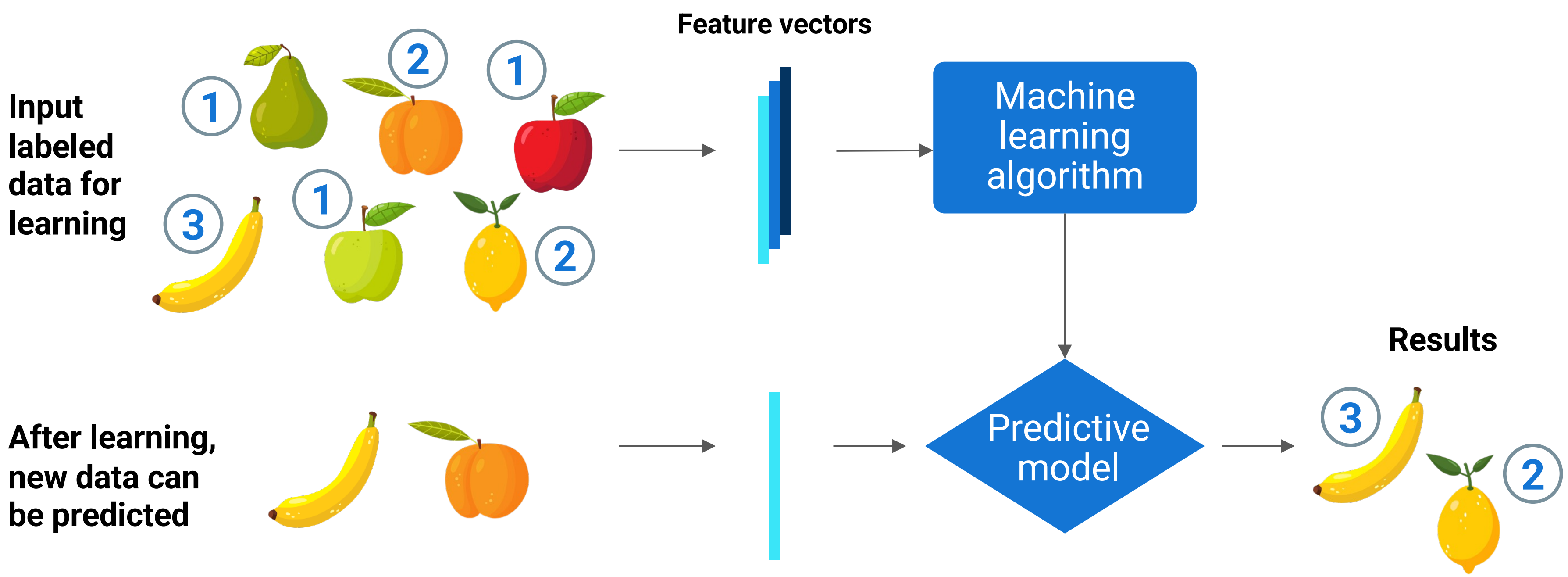


Supervised learning requires us to feed the correct answers to the model.



Introduction to Supervised Learning

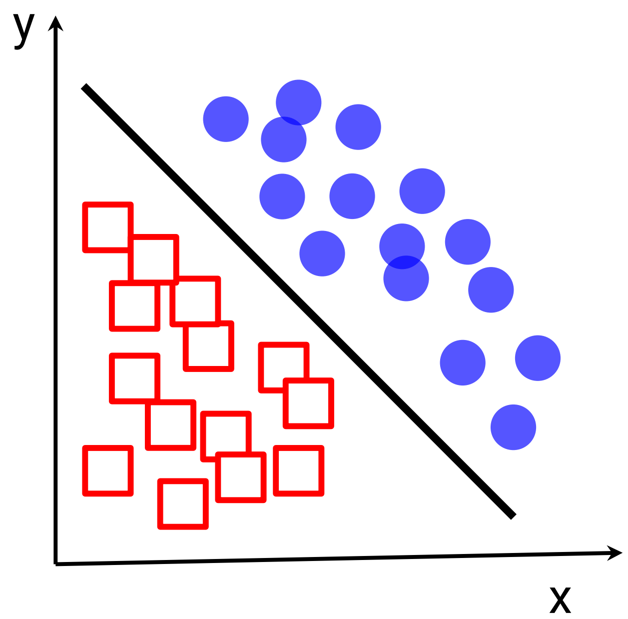
The model learns from the data and the answers. It becomes better at predicting the correct answer as we provide more data.

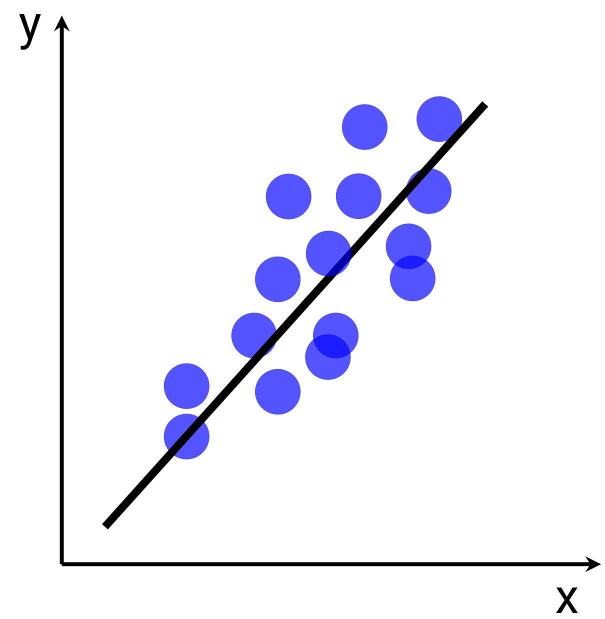


Classification vs. Regression

Classification

Regression

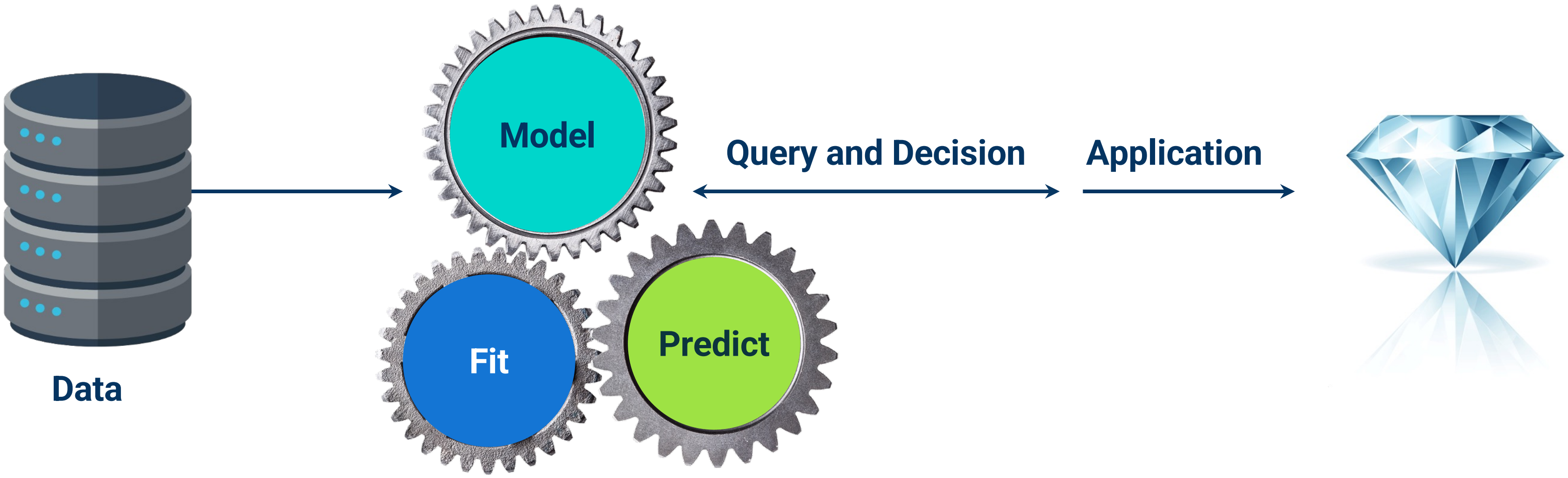




Training and Predicting

Training and Predicting

Regardless of the problem type, in machine learning, we follow a familiar paradigm: Model → Fit (Train) → Predict



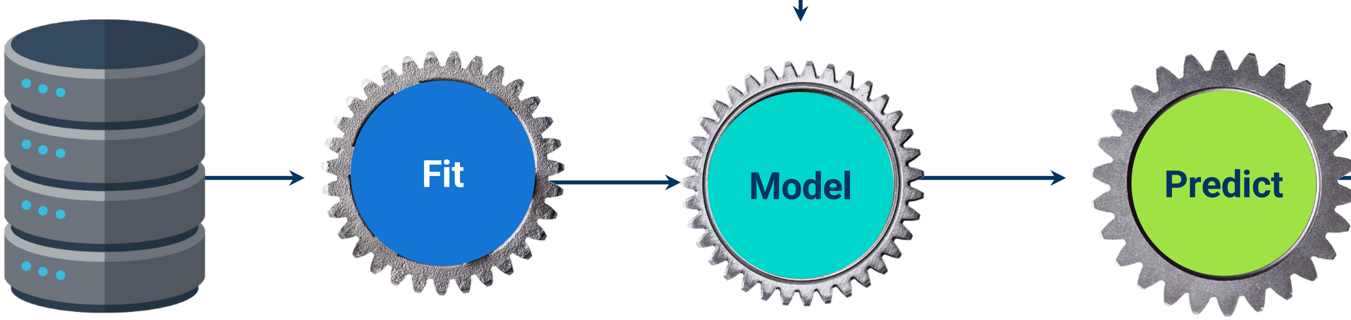
Training and Predicting

Regardless of the problem type, in machine learning, we follow a familiar paradigm: Model → Fit (Train) → Predict

A	B	C	Class
11	16	22	1
10	8	4	2
…	…	…	…

A	B	C	Class
10	15	23	?

Apply

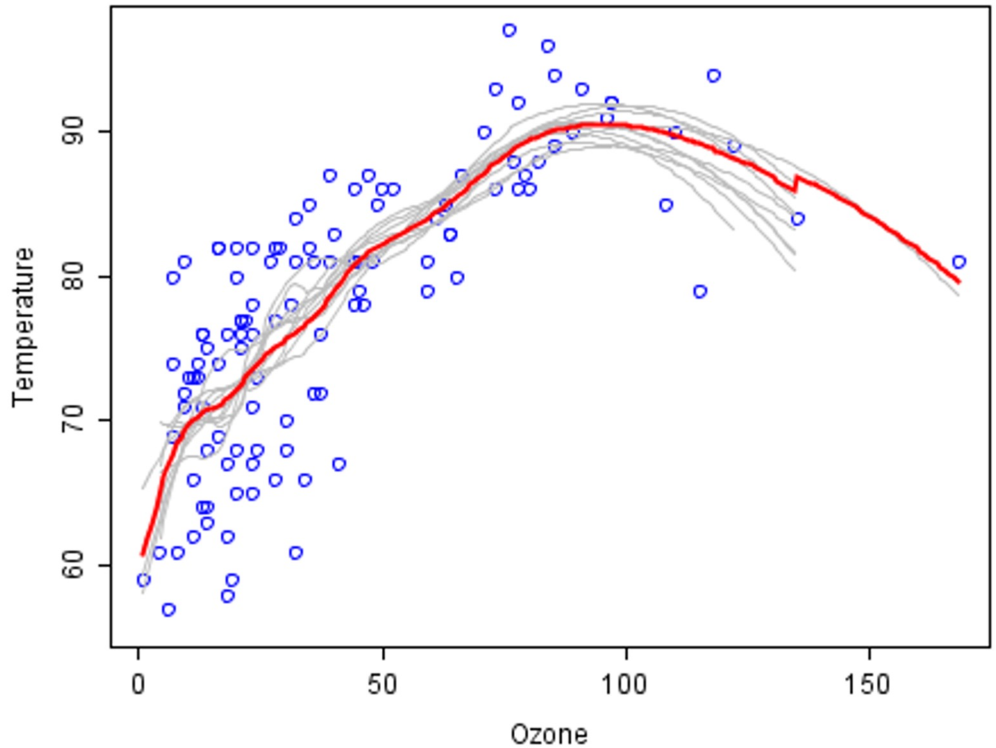


Class
1

Linear Regression

Supervised Learning (Regression)

Regression to predict the location of data points based on old data.



Linear Regression

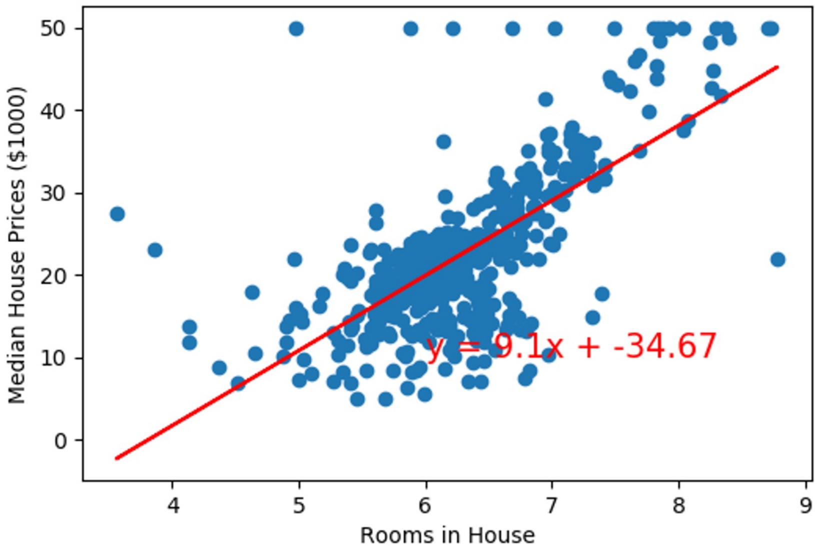
Linear Regression predicts a dependent variable based on values from an independent variable.

There are two basic types:

01 Simple linear regression

02 Multiple linear regression

Both types predict an independent variable using the linear equation.



Linear Regression

The equation of a line: Univariate

$$\underbrace{y}_{\text{Dependent variable}} = \overbrace{m}^{\text{Slope}}\underbrace{x}_{\text{Independent variable}} + \underbrace{b}_{\text{y-intercept}}$$

Linear Regression

The equation of a line: Univariate in Greek!

$$\underbrace{y}_{\text{Dependent variable}} = \underbrace{B_0}_{\text{y-intercept}} + \underbrace{B_1}_{\text{Slope}}\overbrace{x}^{\text{Independent variable}}$$

Linear Regression

The equation of a line: Multivariate

$$\underbrace{y}_{\text{Dependent variable}} = \overbrace{B_0}^{\text{y-intercept}} + \underbrace{B_1}_{\text{Slope}}\overbrace{x_1}^{\text{Independent variable}} + \underbrace{B_2}_{\text{Slope}}\overbrace{x_2}^{\text{Independent variable}} + \cdots + \underbrace{B_n}_{\text{Slope}}\overbrace{x_n}^{\text{Independent variable}}$$

Linear Regression

Linear data trends:

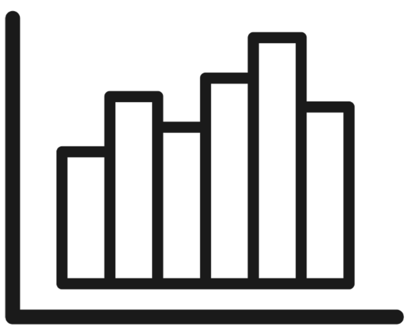
Positive trend	As the independent value `(X)` increases, the dependent value `(y)` increases.
Negative trend	As the independent value `(X)` increases, the dependent value `(y)` decreases.
No trend	As the independent value `(X)` increases, the dependent value `(y)` randomly increases and decreases to the point where there is no clear pattern in the data.

Regression

Regression is a method for predicting **continuous** valued variables.

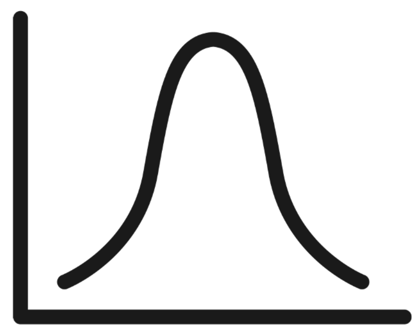
Continuous values

Those that can always be divided into smaller pieces.



Continuous variables

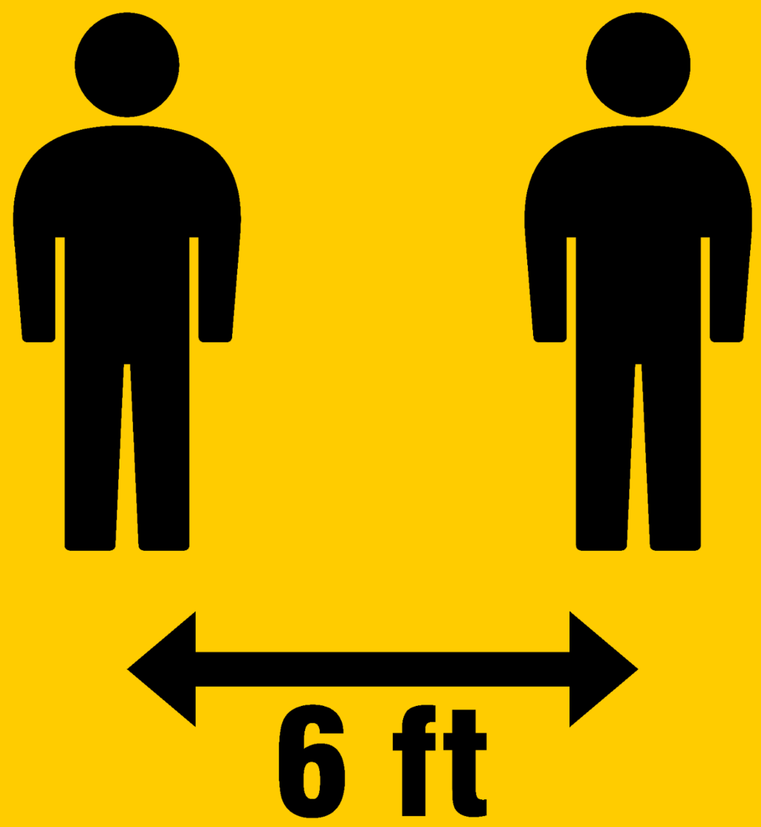
No matter how small, these will have a middle that we can find.



Regression

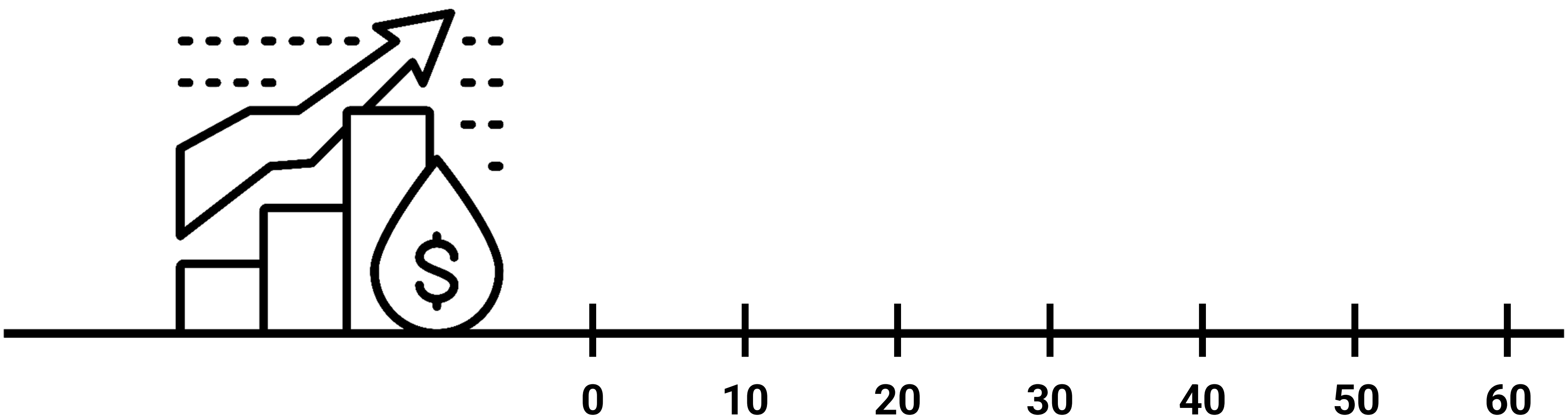
A variable of distance is **continuous**.

We can always find a smaller distance by dividing the current distance by half.



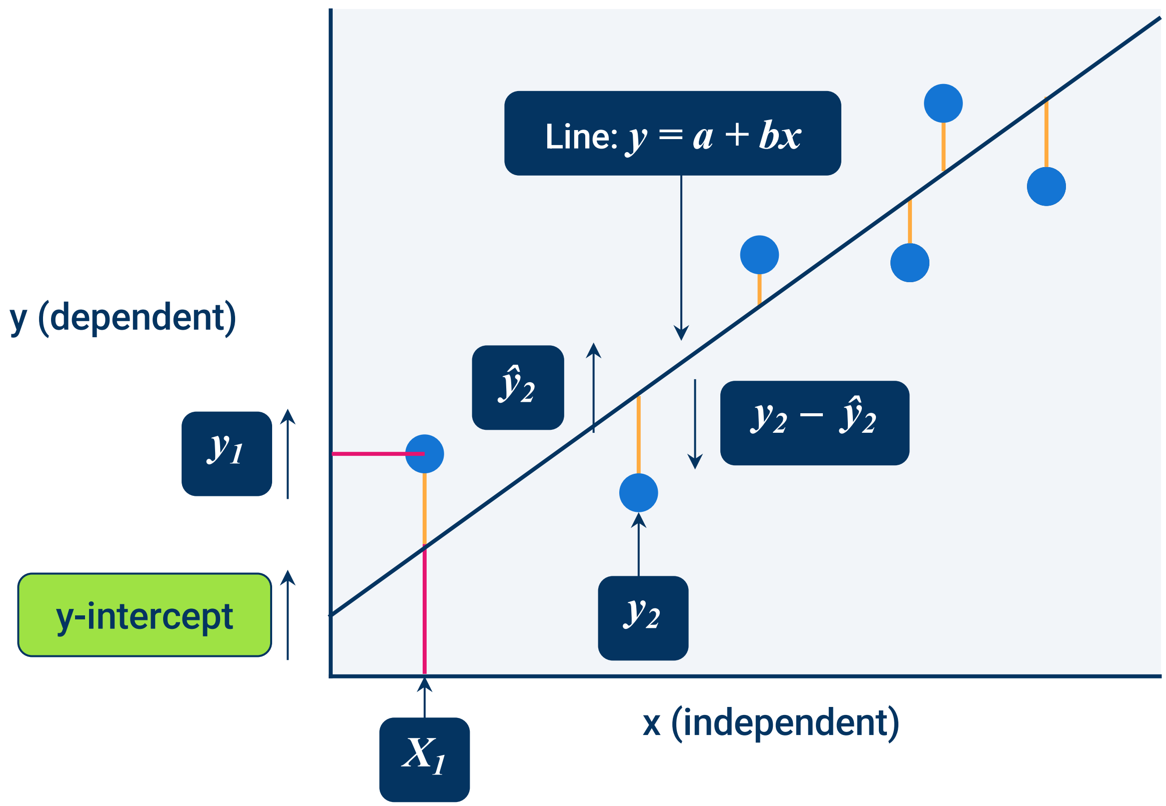
Regression

In finance, prices and rates are usually continuous.



Linear regression is FAST

Univariate Linear Regression Formula



Minimize:

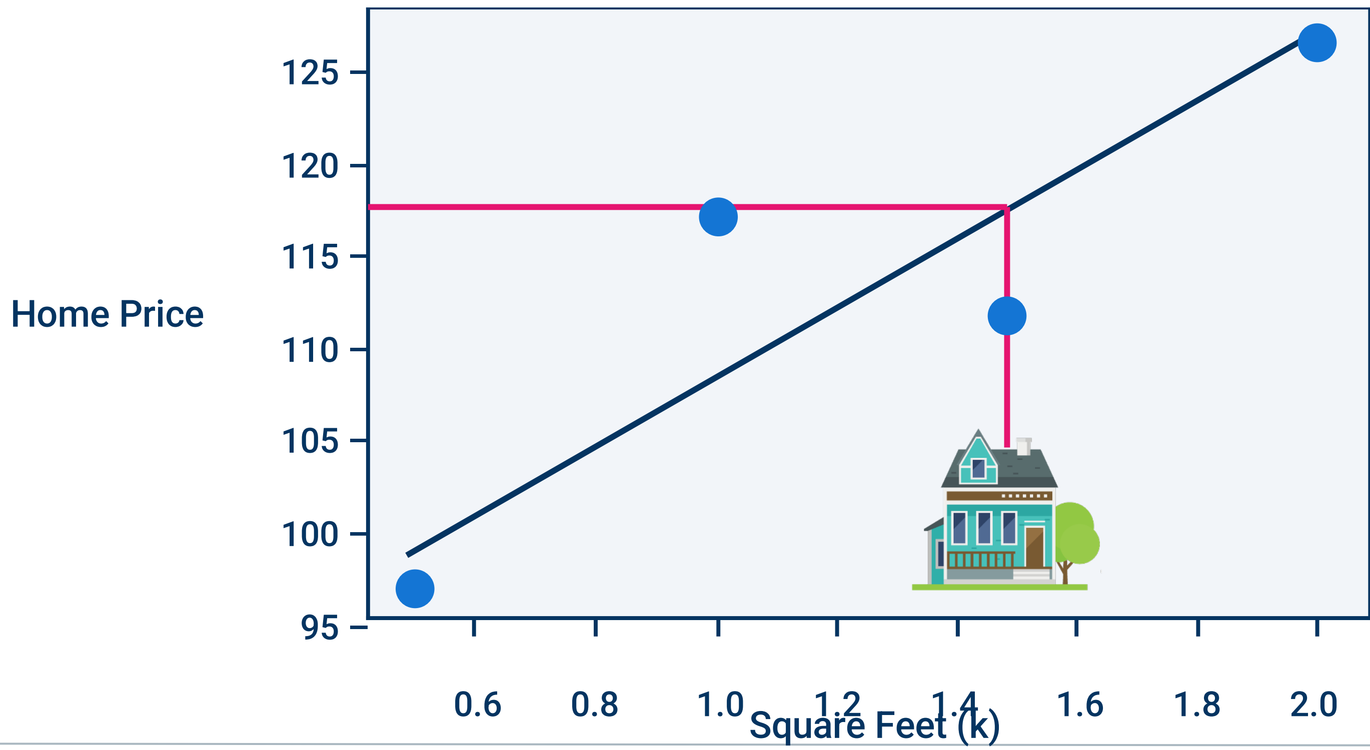
$$\sum^{n} (\, y_i - \hat{y}_i \,)2$$

Least squares method:

$$i = 1$$

Univariate Linear Regression Formula

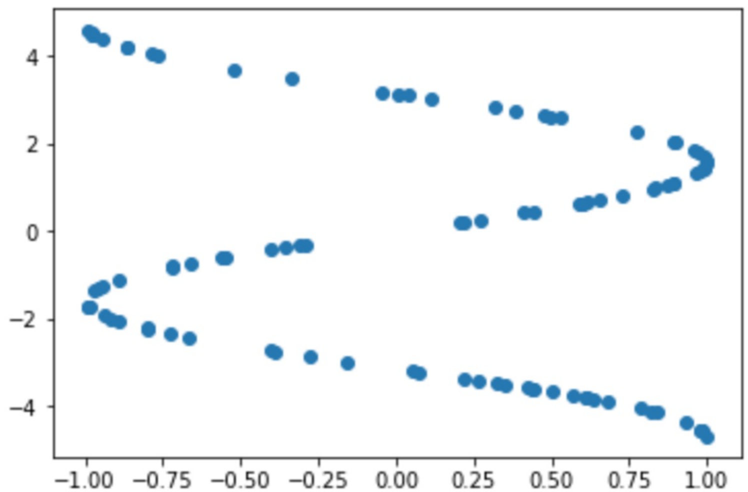
Example using linear regression to predict the home price:



Nonlinear Data

The following image shows a plot of nonlinear data:

```
Out[4]: <matplotlib.collections.PathCollection at 0x11e319c18>
```



Quantifying Regression

Common Scoring Metrics:

R2 (R-Squared)

This is the baseline metric that many ML tools report on score. Higher R2 values signify that the model is “highly predictive.”

An R2 value of >0.90 means that our model roughly accounts for 90% of the variability of the data.

MSE (Mean Squared Error)

This measures the average of the squares of the errors or deviations.

Quantifying Regression

Basic Premise of Validation: Training

We will cut a slice of this data (80%) to build our model.

We'll then use this slice to predict the values for the remaining 20%.

Full Data Set (Historic)

N=1000

# bedrooms	# baths	Sq. feet (k)	Price (k)
2	1	1	200
3	2	1.5	250
...	...	...	...

Training Data Set

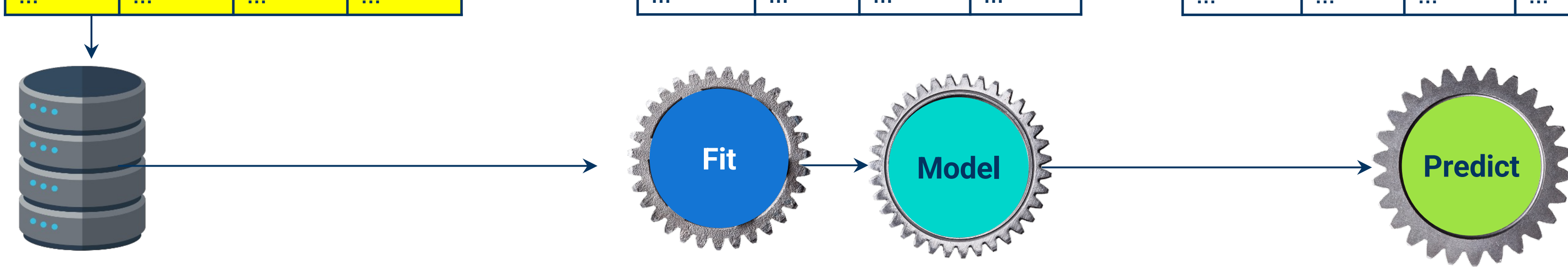
N=800

# bedrooms	# baths	Sq. feet (k)	Price (k)
4	3.5	3.2	450
2	2	1.5	220
...	...	...	...

Testing Data Set

N=200

# bedrooms	# baths	Sq. feet (k)	Price (k)
1	1	.5	60
5	3.5	4.2	780
...	...	...	...



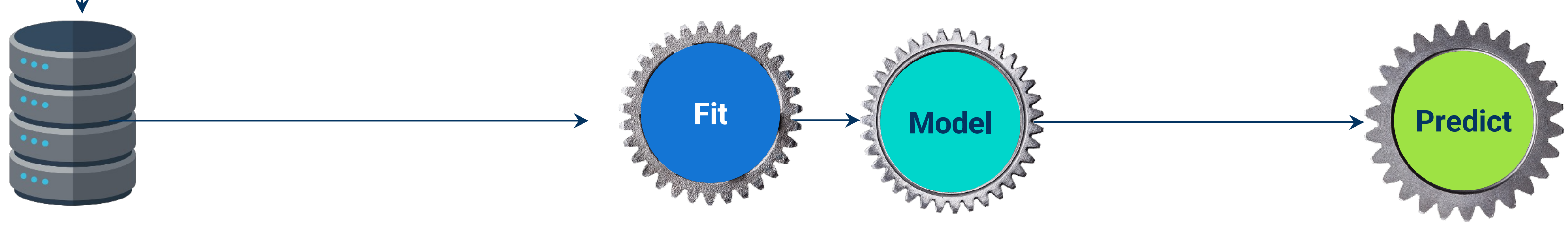
Quantifying Regression

Basic Premise of Validation Using Training/Testing Data: We use the training data to fit the model to the data. This is the training step where we build a model that can predict our output (home price) for a given set of features (# bedrooms, # baths, square feet).
Once the model is trained, we can use the model to make predictions.

Full Data Set (Historic)			
N=1000			
# bedrooms	# baths	Sq. feet (k)	Price (k)
2	1	1	200
3	2	1.5	250
…	…	…	…

Training Data Set			
N=800			
# bedrooms	# baths	Sq. feet (k)	Price (k)
4	3.5	3.2	450
2	2	1.5	220
…	…	…	…

Testing Data Set			
N=200			
# bedrooms	# baths	Sq. feet (k)	Price (k)
1	1	.5	60
5	3.5	4.2	780
…	…	…	…



Quantifying Regression

Basic Premise of Validation: We use the test data to make new home price predictions. We can then compare the home price of our prediction vs. the actual price. Based roughly on how often we are “correct,” we get a score for the model as a whole. If the model scores well, we can trust it for future use. We train the model on the training data and score the model based on data that it has never seen before (test data).

Full Data Set (Historic)

N=1000

# bedrooms	# baths	Sq. feet (k)	Price (k)
2	1	1	200
3	2	1.5	250
…	…	…	…

Training Data Set

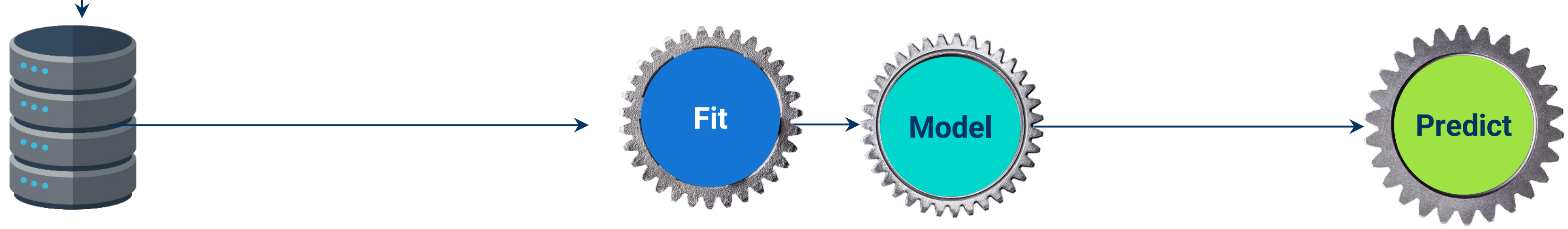
N=800

# bedrooms	# baths	Sq. feet (k)	Price (k)
4	3.5	3.2	450
2	2	1.5	220
…	…	…	…

Testing Data Set

N=200

# bedrooms	# baths	Sq. feet (k)	Price (k)
1	1	.5	60
5	3.5	4.2	780
…	…	…	…

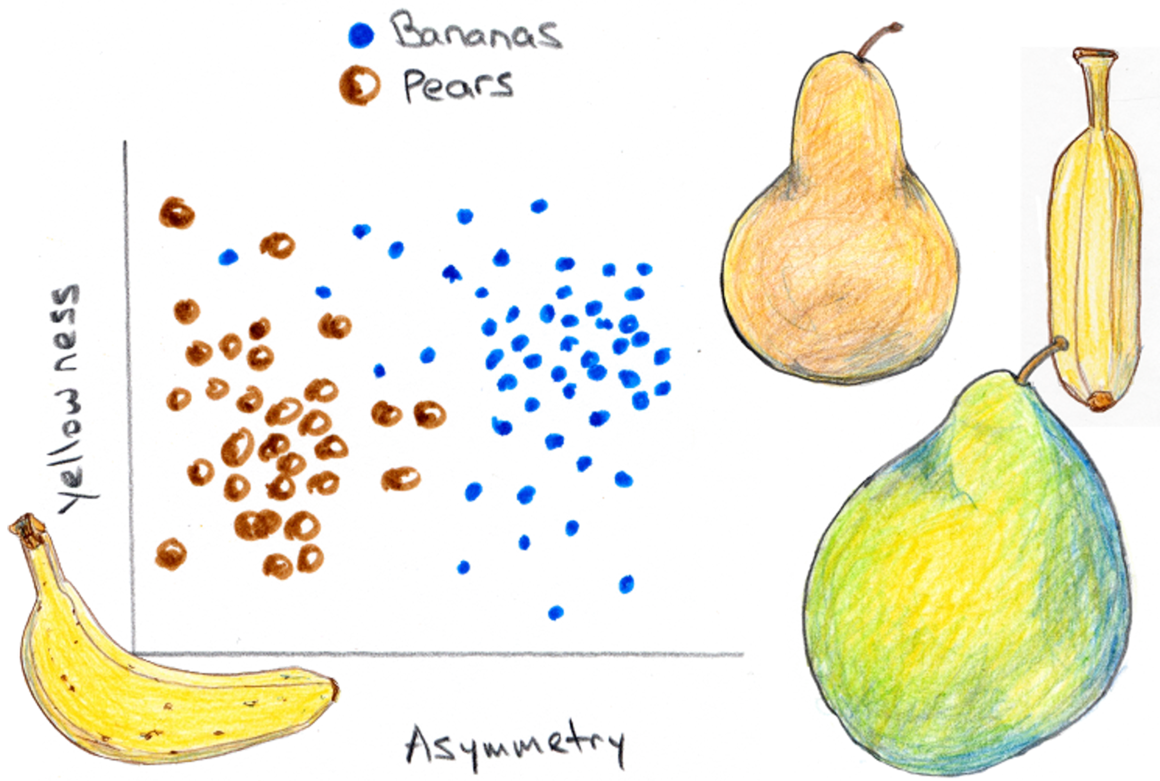


What is classification?

Classification is a method to predict discrete valued variables.

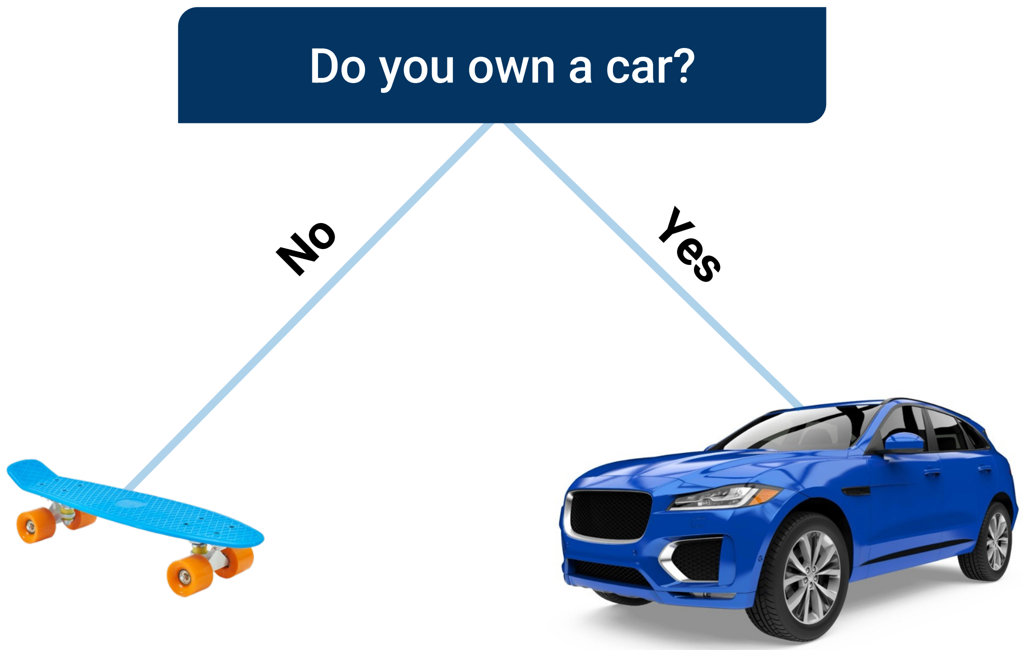
A discrete variable has no middle, and its values cannot be divided.

Supervised Learning (Classification)



Classification

Consider a loan application that asks:



- The possible answers are yes or no.
- You either own a car or you don't.
- There is no middle value, so this type of variable, such as `car_ownership`, would be discrete.

Classification

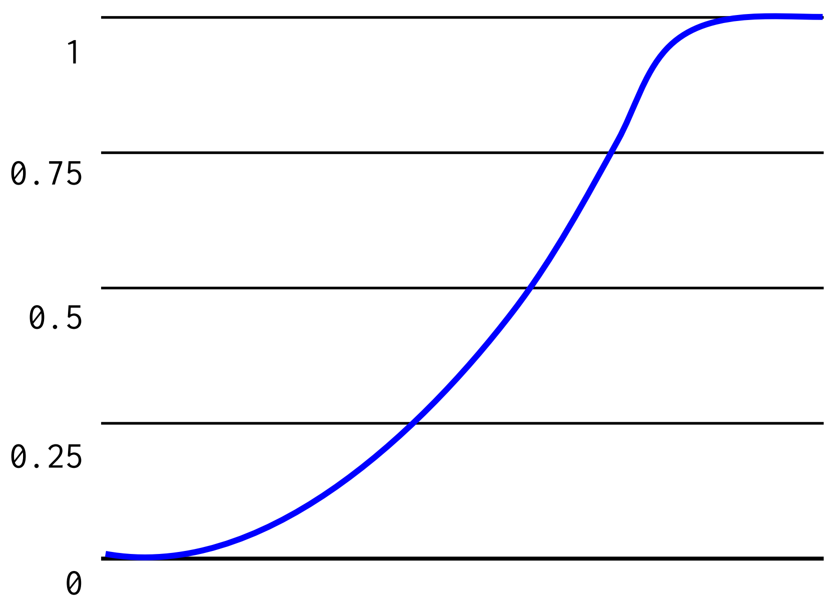
Fintech analysts use classification to draw categorical conclusions about data.

Instead of forecasting quantitative numbers, classification uses a binary (true-positive / true-negative) approach to predict membership in a category (i.e., will the outcome be of type A or type B).

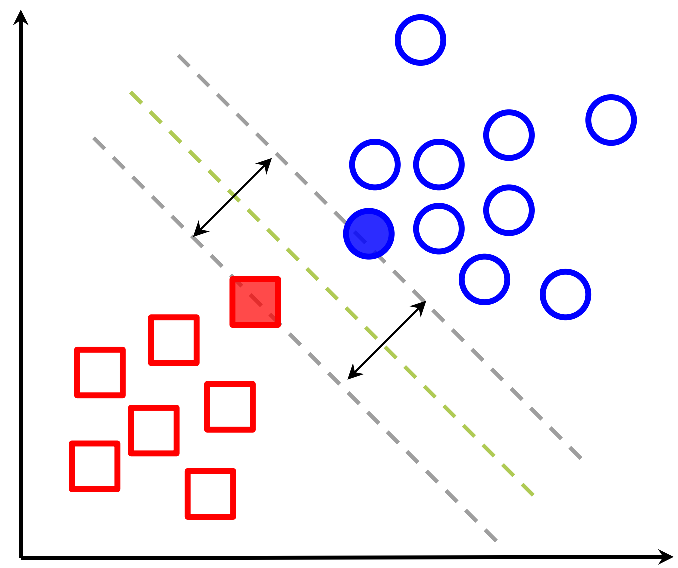
Qualified loan applicant	VS.	Unqualified loan applicant
Creditworthy		Credit risk
0		1

Classification

Today you'll learn to perform classification using **logistic regression**.



In the next class, you'll discover other tools for classification including **support vector machines**, decision trees, and random forests.



Classification

Classification models have greatly improved the ability for organizations to properly classify applicants, predict market decline, and classify fraudulent transactions or suspicious activity.

- Most large financial institutions use some form of machine learning to monitor and predict fraudulent activities.
- This is how banks know when to flag and decline transactions due to suspicion of fraud.

Classification

FICO credit scoring currently uses a classification model for their cognitive fraud analytics platform.

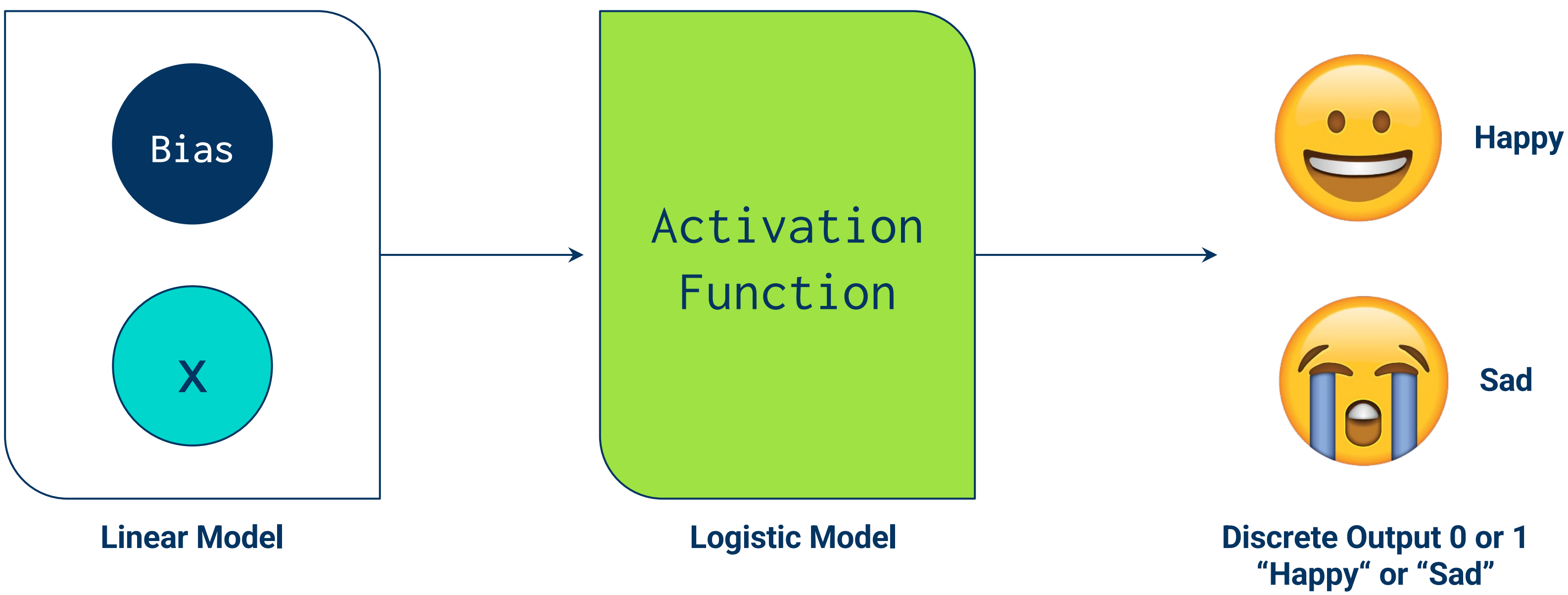
- Classification models have allowed the financial industry to become more proactive.
- Supervised learning algorithms can predict outcomes with a high degree of accuracy, which allows for more effective and efficient mitigation.

Logistic Regression

Logistic regression is a classification algorithm used to predict a discrete set of classes or categories (e.g., Yes/No, Young/Old, Happy/Sad).

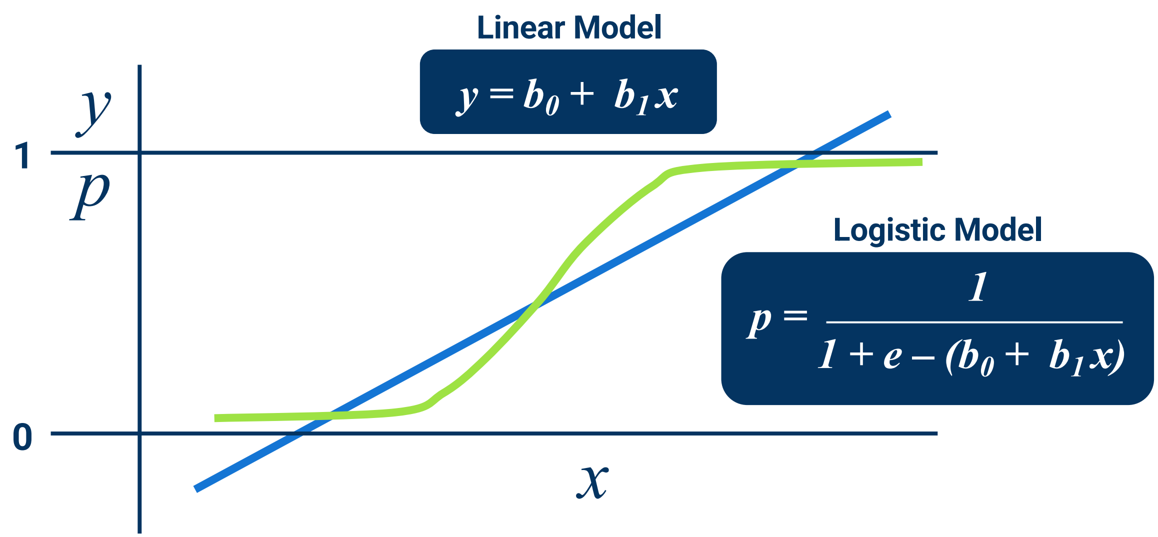
Logistic Regression

Logistic regression is a classification algorithm used to predict a discrete set of classes or categories (e.g., Yes/No, Young/Old, Happy/Sad).



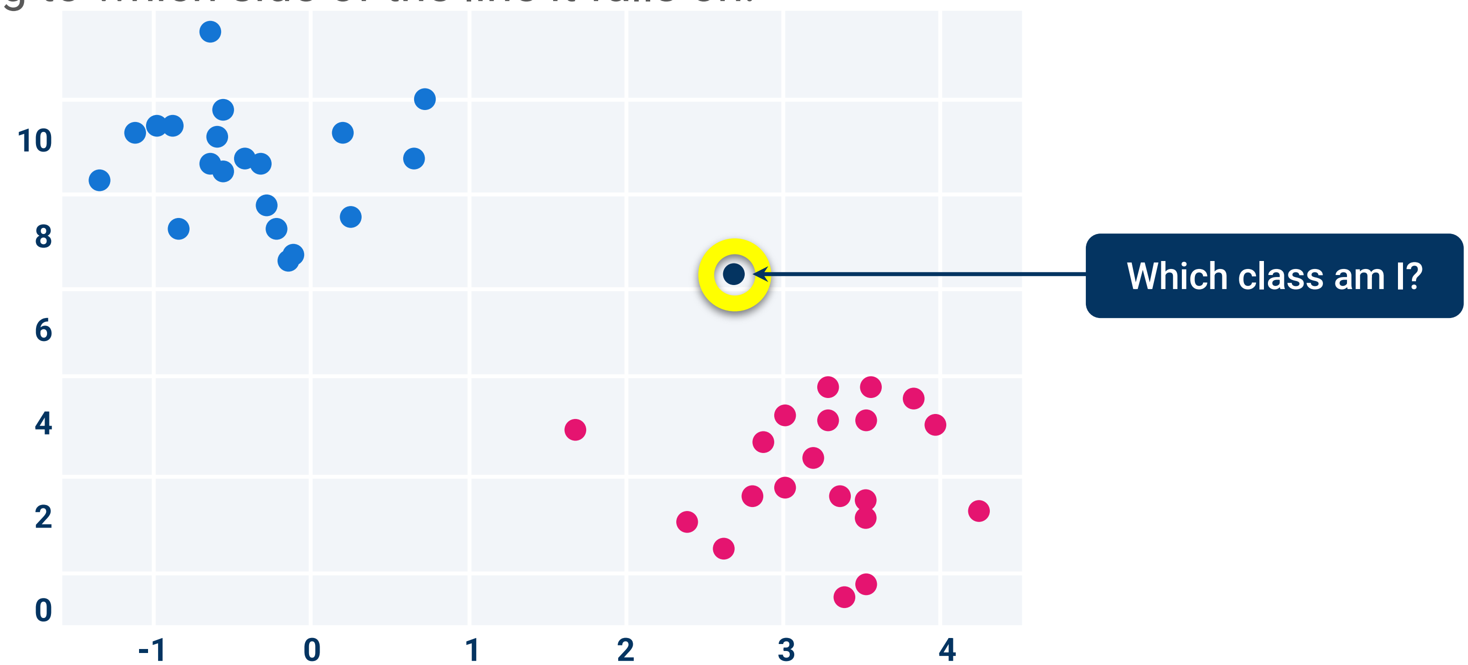
Logistic Regression

Unlike linear regression, which outputs continuous numerical values (for example, age), logistic regression applies an activation function, such as the sigmoid function, to return a probability value of 0 or 1. This can then be mapped to a discrete class like "Young" or "Old."



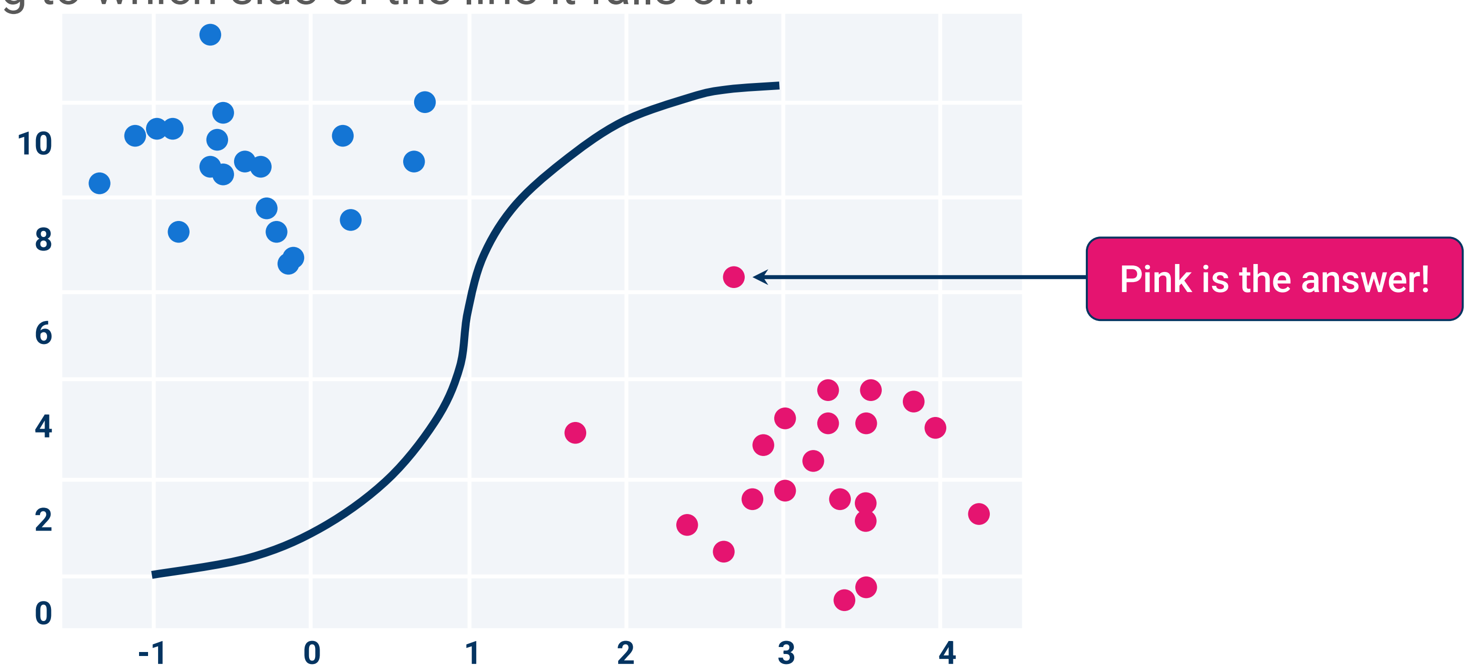
Logistic Regression

Applying logistic regression gives us a line on the plot that separates the two classes. Now, we can predict which class a new data point should belong to—according to which side of the line it falls on.



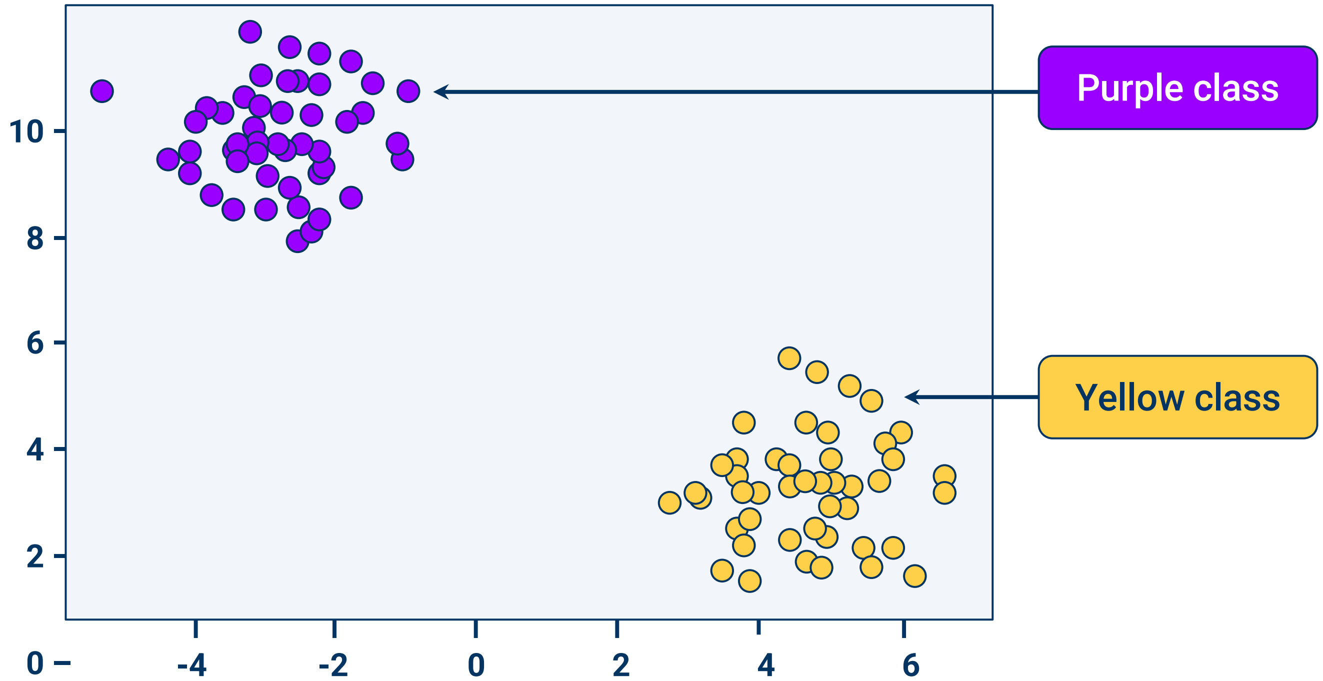
Logistic Regression

Applying logistic regression gives us a line on the plot that separates the two classes. Now, we can predict which class a new data point should belong to—according to which side of the line it falls on.



Logistic Regression

Apply logistic regression to determine if new data points belong to the purple group or the yellow group.



Confusion Matrix

A Confusion Matrix compares the predicted values from a model against the actual values. The entries of the confusion matrix are the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

	Predicted True	**Predicted False**
Actually True	113 (True Positives)	12 (False Negatives)
Actually False	31 (False Positives)	36 (True Negatives)

Confusion Matrix

We can calculate measures, such as the accuracy of a model, from the values in the confusion matrix.

$$\text{Accuracy} = TP + TN / (TP + TN + FP + FN)$$

We can also calculate other measures that gives us more information, such as:

- Precision
- Sensitivity
- F1 Score
- Positive Predictive Value
- Negative Predictive Value
- Threat Score
- False Omission Rate

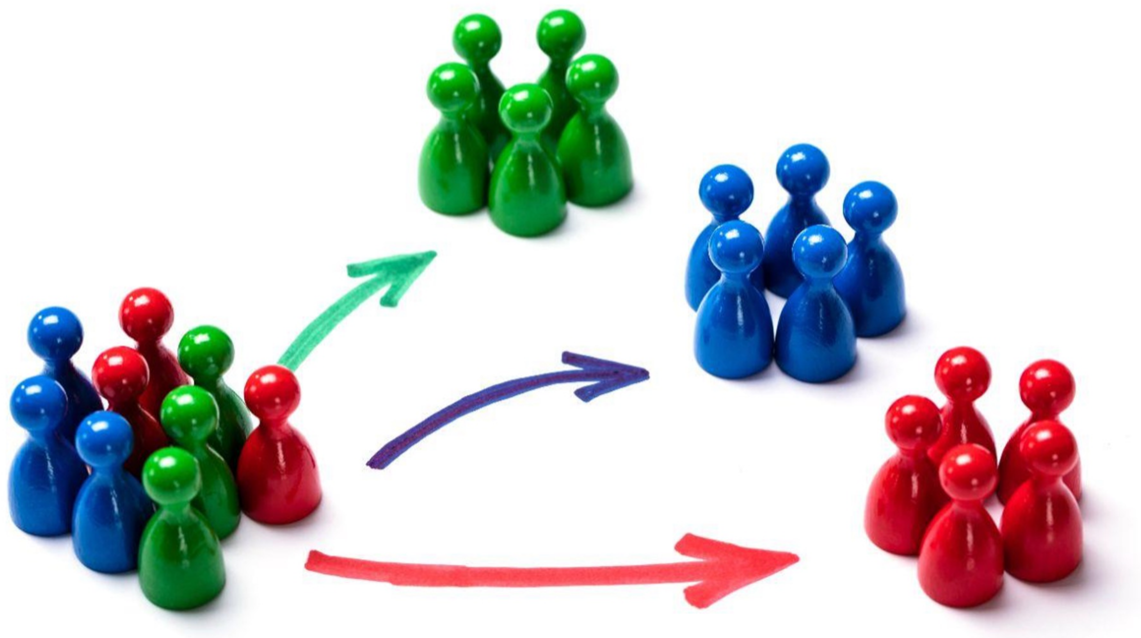
Unsupervised learning allows us to cluster data to find hidden or unknown patterns.

Customer Segmentation

Customer segmentation is one of the most popular applications of unsupervised learning. It categorizes customers based on their demographic and behavioral traits.

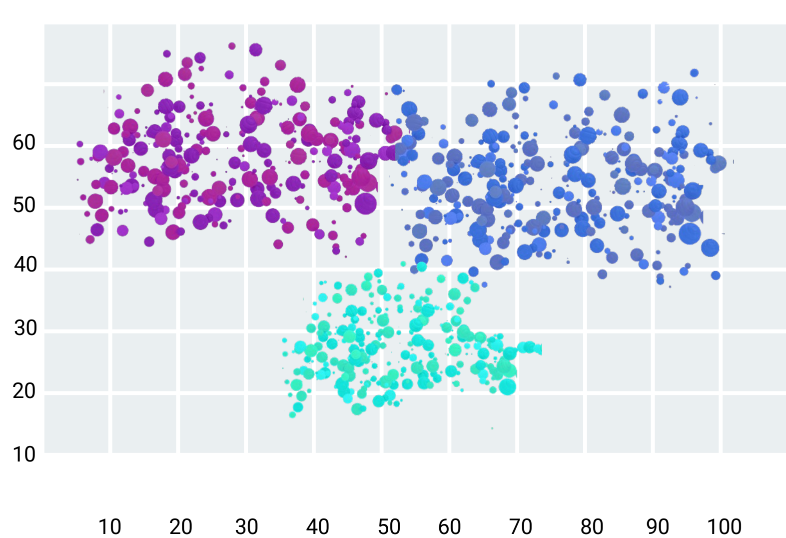
We can group customer-based similarities, such as:

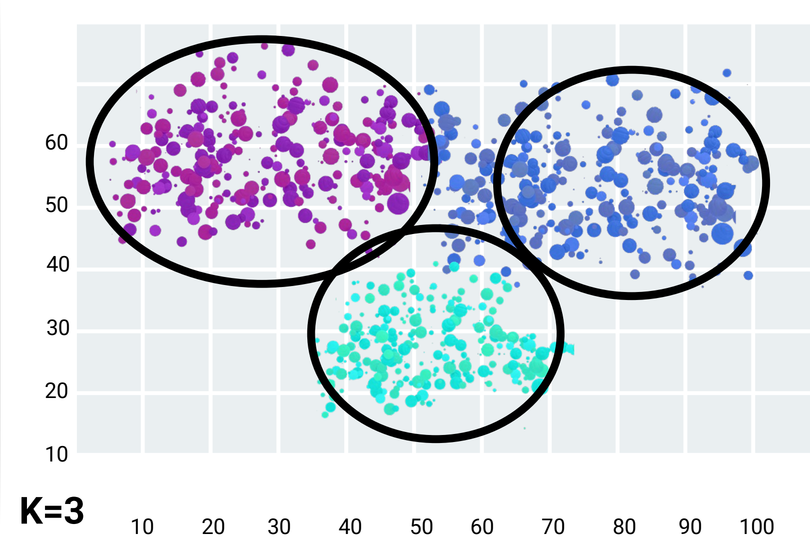
- **Customer needs**
  (e.g., group A needs groceries while group B needs home decor)
- **Responses to online marketing channels**
- **Buying habits**
  (eg. best day for buying, weekly spend)



Clustering

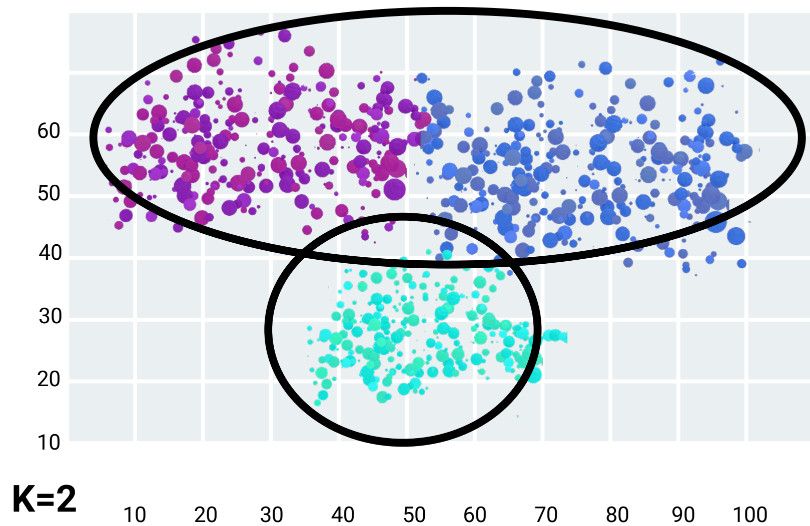
In this clustering problem, we expect our algorithm to group data points based on their mutual similarities of features.





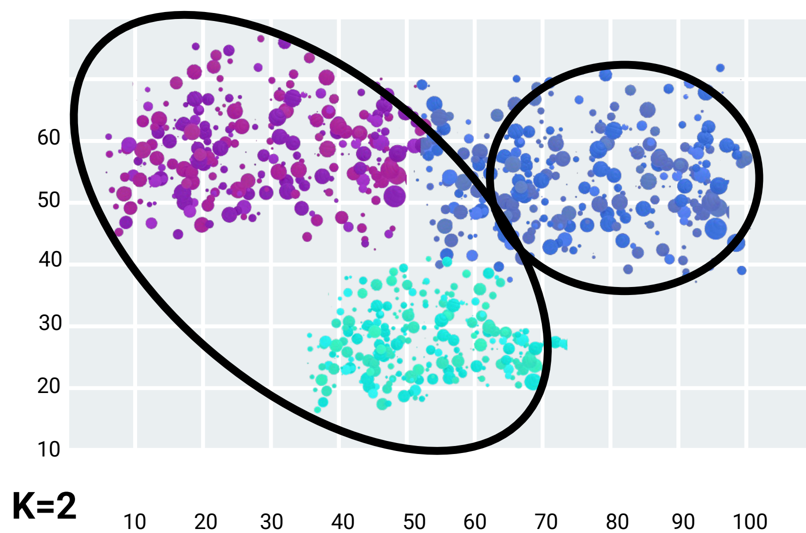
Clustering

But the problem is more complex.



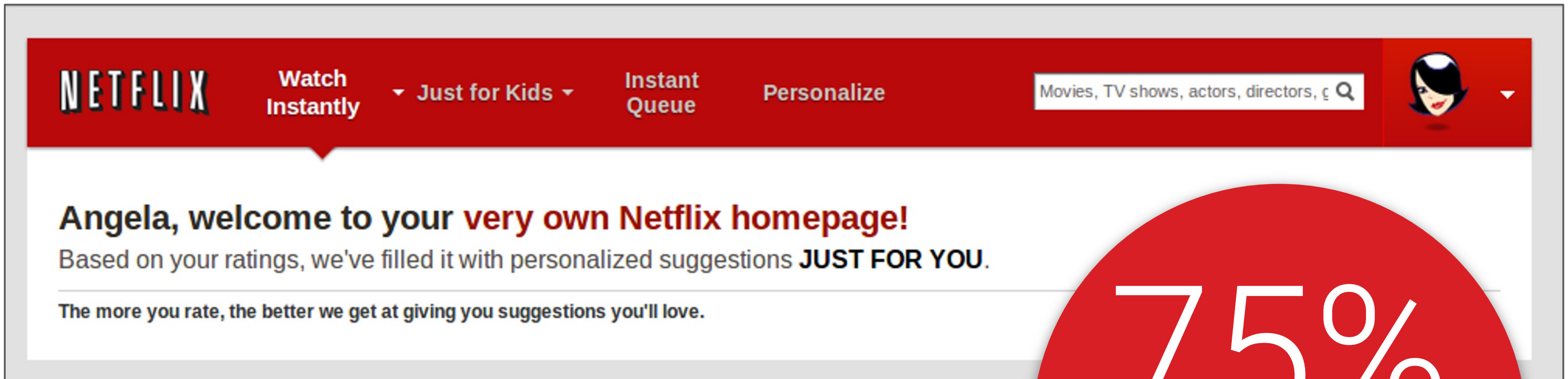
Clustering

Perhaps the clusters are not where we think they are.



Customer Segmentation

Customer segmentation is driving revenue in leading companies like Netflix:



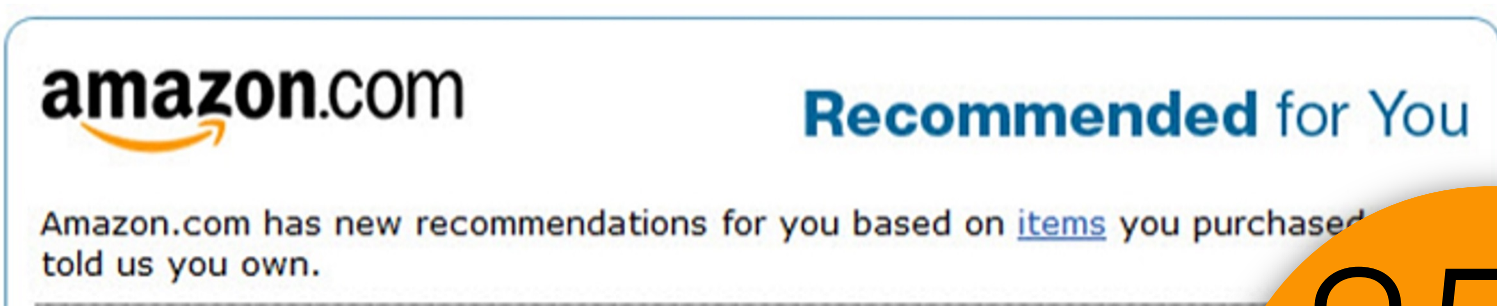
75%

of viewer activity is estimated to be driven by recommendation.

Netflix's recommendation system saves the company an estimated $1 billion per year through reduced churn.

Customer Segmentation

Customer segmentation is driving revenue in leading companies like Amazon.



35%

of sales are estimated to be generated through the recommendation engine.

Data Preprocessing

Real life data almost always needs to be processed before it can be used in a machine learning algorithm

Preprocessing Data

Two major preprocessing steps are converting categorical data and scaling:

Converting Categorical Data

Categorical data is non-numeric data, like the day of the week or a person's education level, and needs to be converted to numeric data.

Scaling

Some machine learning algorithms are sensitive to large data values, so features need to be scaled to standardized ranges.

One-Hot Encoding and Label Encoding

Label Encoding

Label Encoding turns categorical variables into a series of integers, for example, “Sunday” becomes 0, “Monday” becomes 1, “Tuesday” becomes 2, and so on.

It can cause problems though, because the difference between Saturday and Sunday in our previous example is -6, but the difference for other consecutive days is +1.

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
0	1	2	3	4	5	6

One-Hot Encoding

One-Hot Encoding creates new “dummy” features for each category with 0 and 1 as Boolean values.

So the Weekday feature becomes 6 new features:

`isSunday`	`isMonday`	`isTuesday`	`isWednesday`	`isThursday`	`isFriday`	

Why is there no `isSaturday`?

One-Hot Encoding

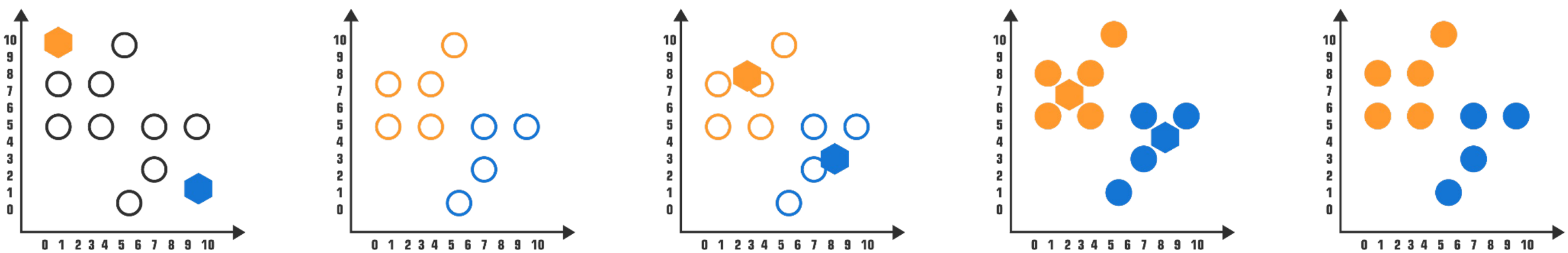
`isSaturday` can be reconstructed from the other six. This means `isSaturday` is **collinear** with the other dummy features. Including `isSaturday` is an example of "the dummy trap," which can cause errors in the machine learning model.

`isSunday`	`isMonday`	`isTuesday`	`isWednesday`	`isThursday`	`isFriday`	

For each observation, only one of the new features will have a value of 1. This is why we call it "**one-hot encoding.**"

Preparing Data by Normalizing It

We can optimize data clustering by selecting the best value for k. The K-means algorithm is useful for grouping and understanding financial data.



Randomly select K-clusters

Each object assigned to similar centroid randomly

Clusters centers updated depending on renewed cluster mean

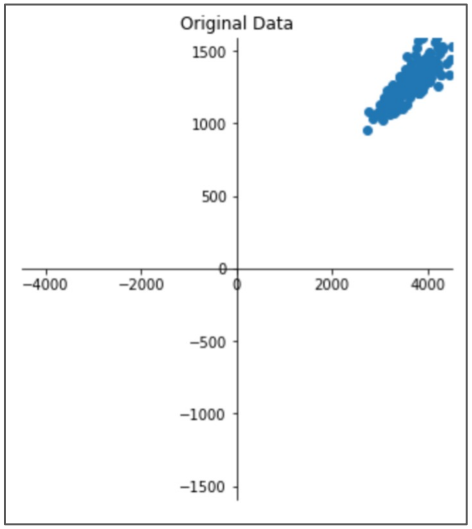
Re-assign data points; update cluster centers

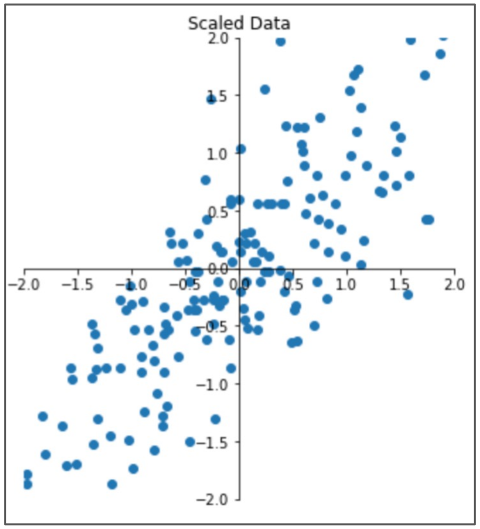
Re-assign data points

Scaling/Normalization

We want all features to be shifted to similar numeric scales so that the magnitude of one feature doesn't bias the model during training.

`StandardScaler`	Scales data to have a mean of 0 and variance of 1. You would use `StandardScaler` when you do not have complete knowledge of your data.

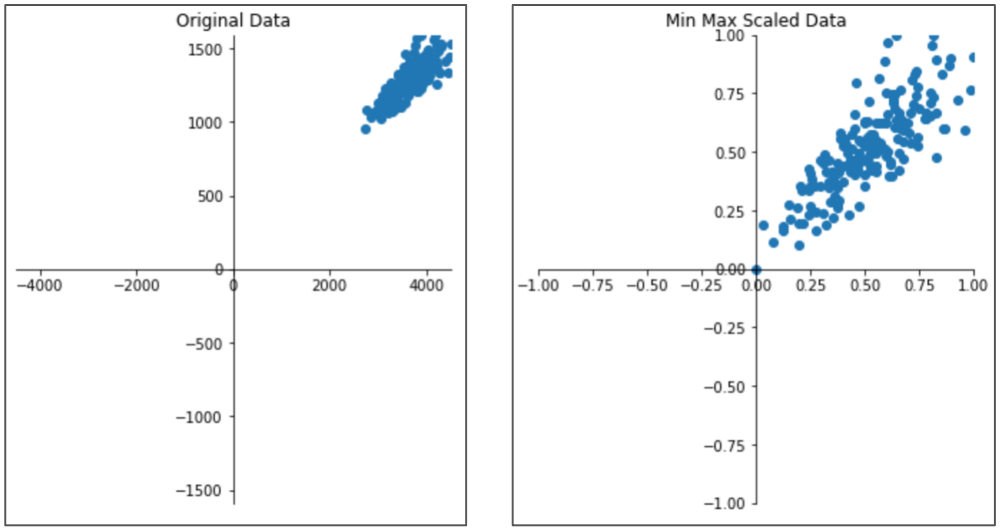




Scaling/Normalization

We want all features to be shifted to similar numeric scales so that the magnitude of one feature doesn't bias the model during training.

`MinMaxScaler`	`MinMaxScaler` scales feature data to a minimum of 0 and a maximum of 1.



Normalizing Data

Remember, the K-means algorithm requires all the columns in a DataFrame to have numeric values.

- We should also ensure that the numeric values have the same scale.
- This prevents K-means from putting too much weight on any single variable.

Numeric Data Before Normalizing

eps	times_sales	total_assets	total_debt
2.61	63.73	222822.05	46244.82
0.12	17.55	234.42	0.00
7.96	44.14	239.78	15.24
-21.25	109.27	16872.89	0.00
62.48	387.85	156035.77	41128.51

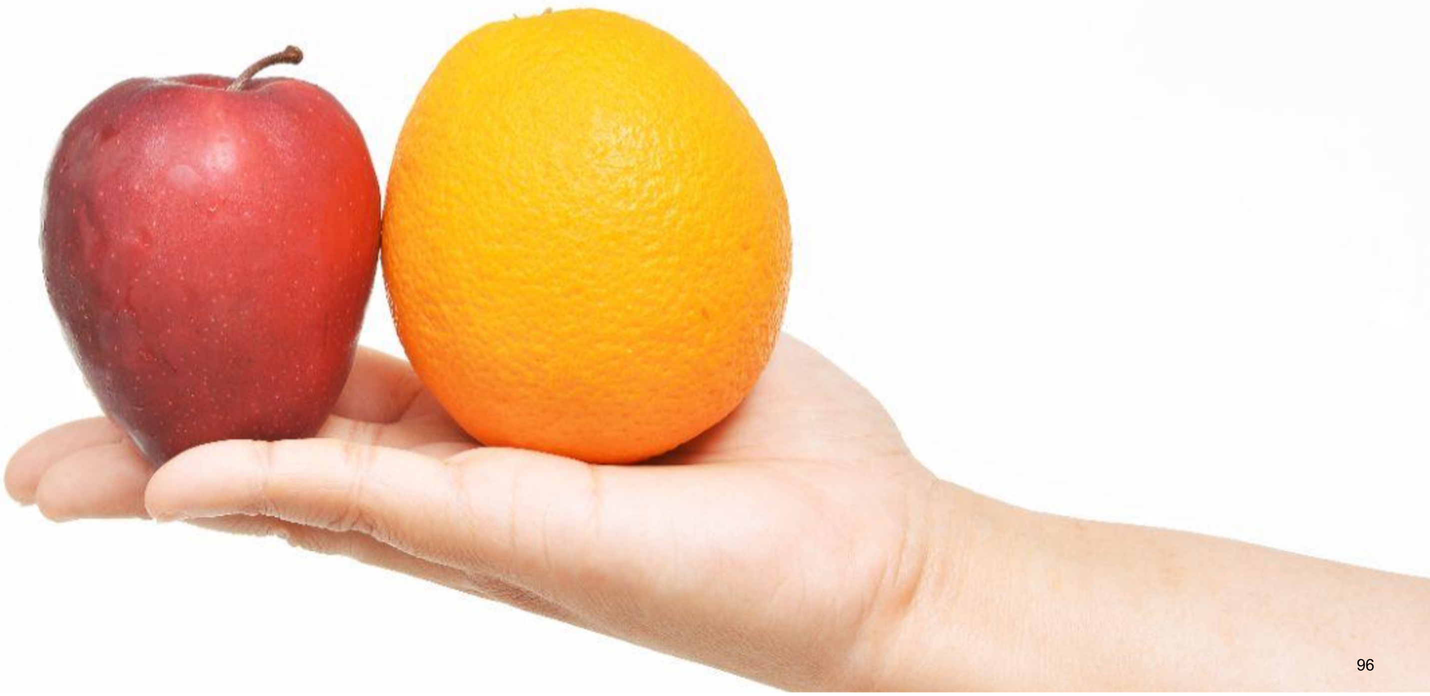
The Same Data After Normalizing

eps	times_sales	total_assets	total_debt
-0.0575	-0.0797	-0.1134	-0.0864
-0.0570	0.0795	-0.1136	0.0864
-0.0594	-0.0796	-0.1136	-0.0864
-0.0567	-0.0770	0.1135	-0.0862
0.0484	0.2537	-0.0961	-0.0836

Normalizing Data

When we **normalize data**, we eliminate the measurement units and scale the numeric values to a similar scale.

We can then compare data of **differing natures**.



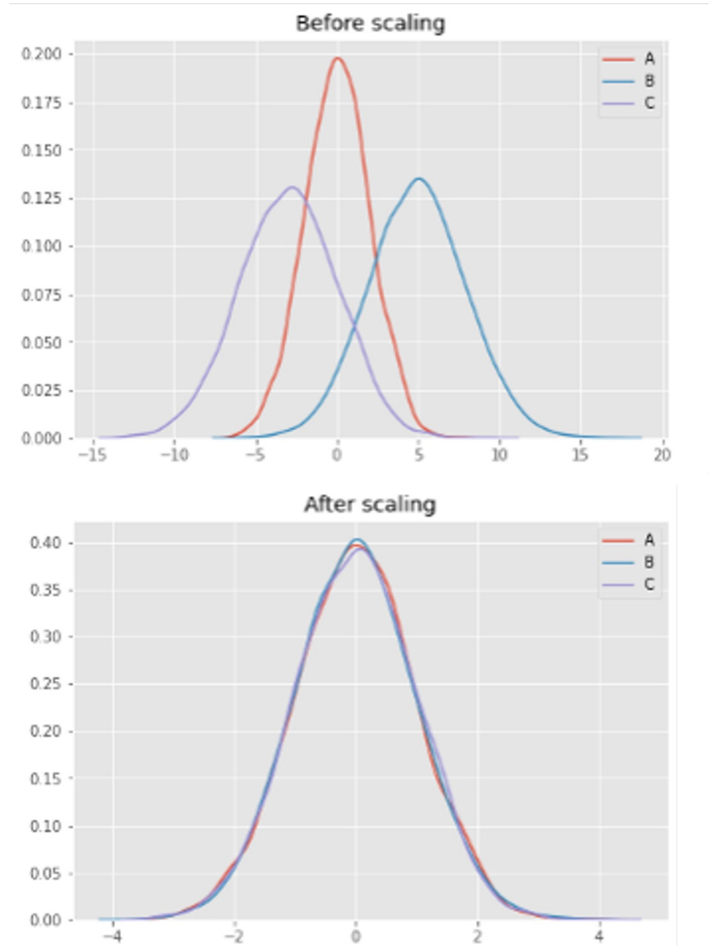
Normalizing Data

The most common way to normalize data is to apply **standard scaling**, which is a method of centering values around the mean.

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

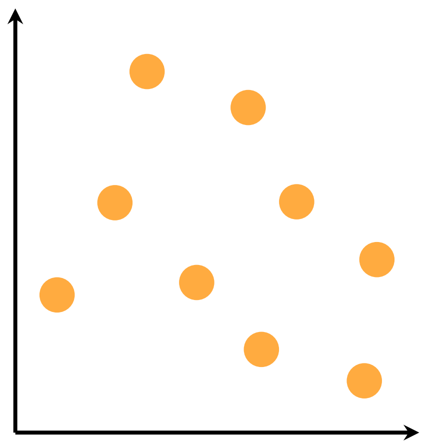
σ = Standard deviation



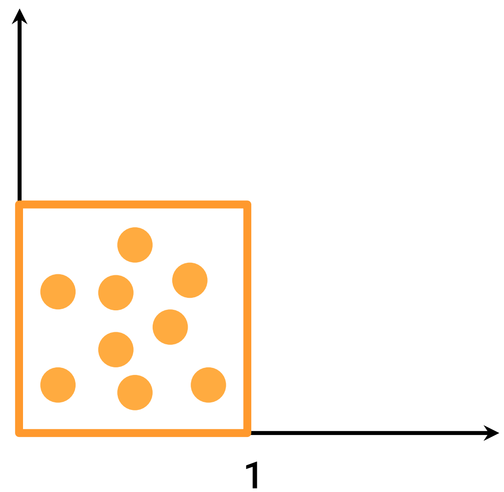
Normalizing Data

Data standardization, or **data normalization**, is a common practice in the data preprocessing steps that occur before training a machine learning model.

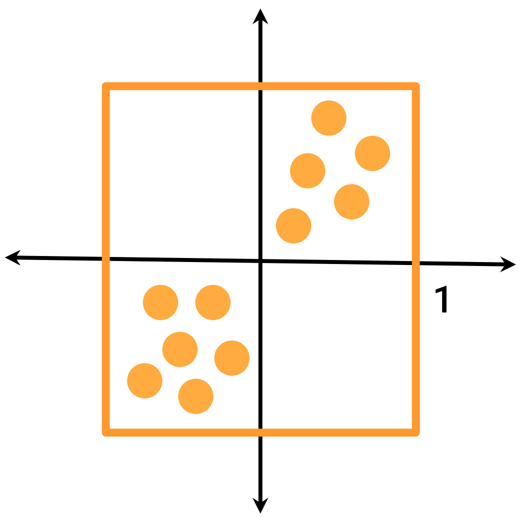
Actual data



After normalizing



After standardization



KNN: Logistic Regression Friendly Neighbor

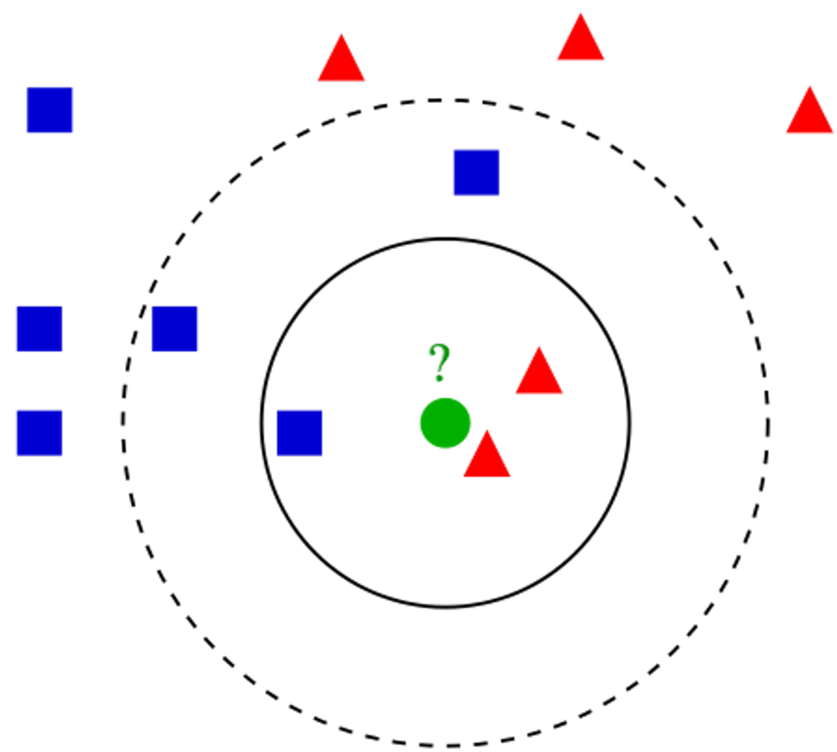
One of the most popular and flexible supervised machine learning models is the k-nearest neighbor model.

The **k-nearest neighbor** compares known data points to determine the classification of a novel data point.

KNN: Logistic Regression Friendly Neighbor

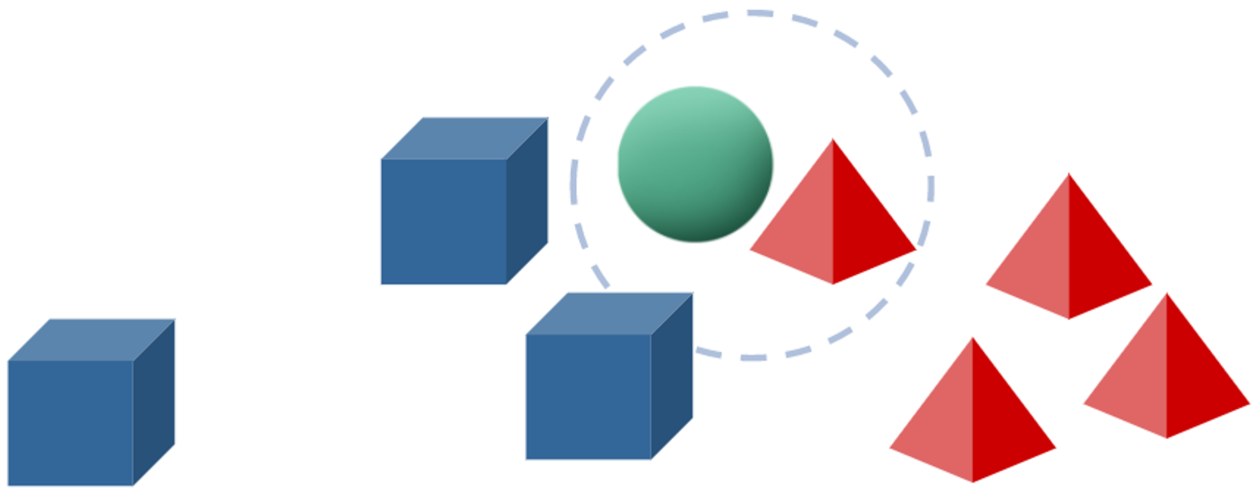
We set the k-value parameter to tell the algorithm to find the k-number of closest known data points.

Then, the algorithm determines what the majority of surrounding data points are classified to determine the class of the new data point.



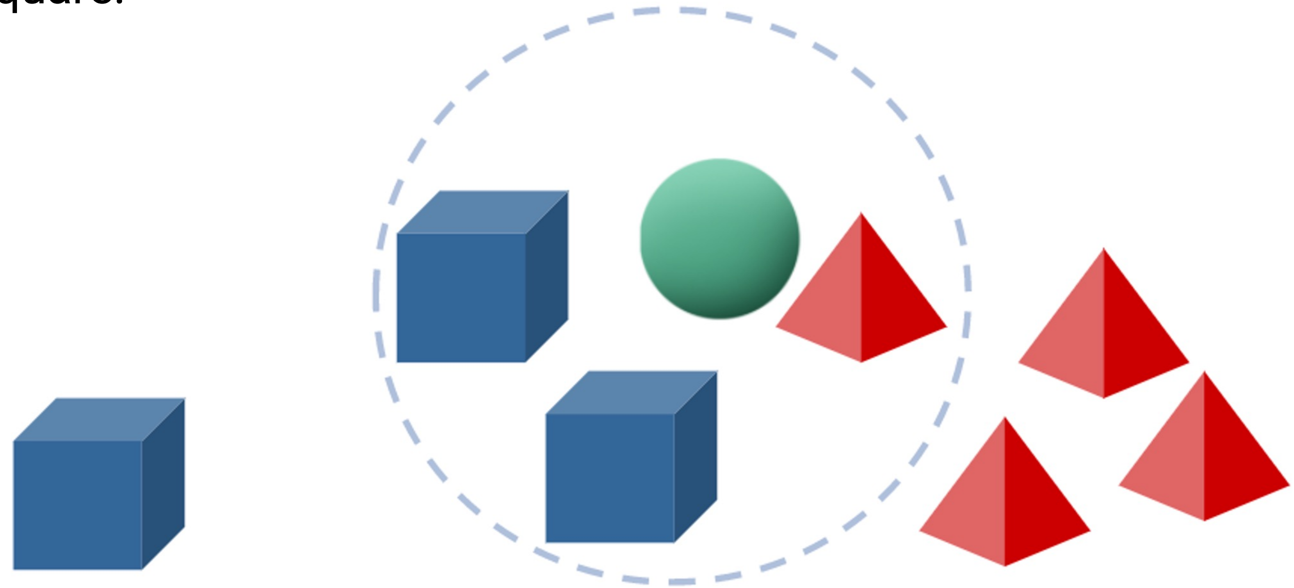
KNN: Logistic Regression Friendly Neighbor

Consider the example we wanted to classify a new data point (green circle) from known square and triangle data points. If **k = 1**, then the algorithm classifies our new data point to whatever is the closest known single neighbor. In this case our green circle would be classified as a red triangle.



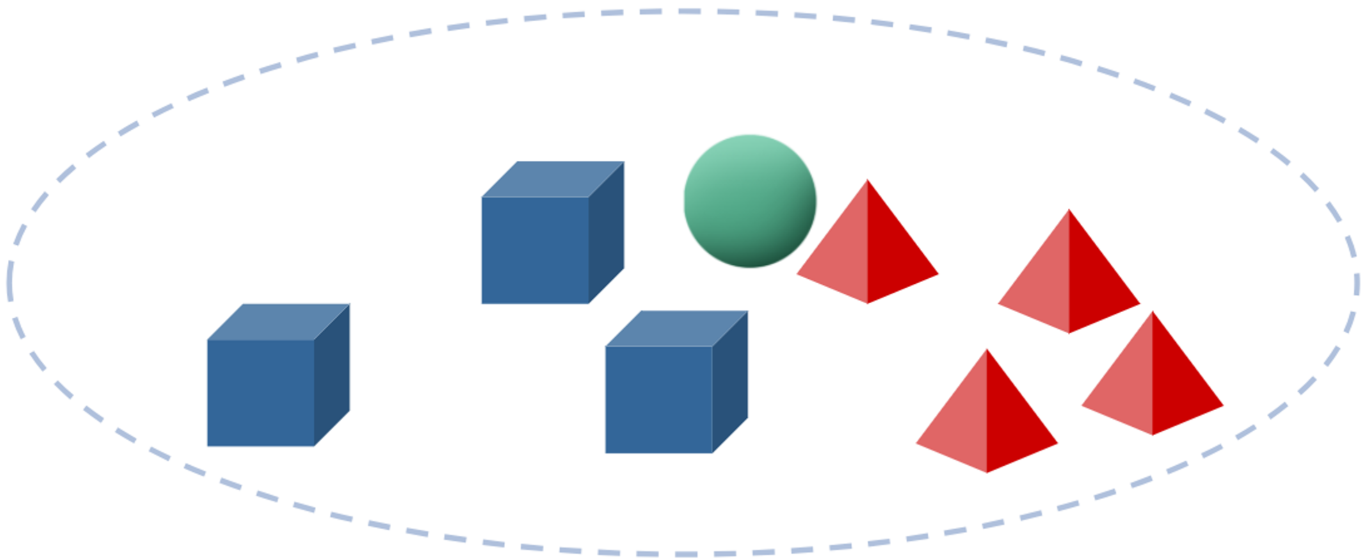
KNN: Logistic Regression Friendly Neighbor

As the number of k-nearest neighbors increase, the distribution of nearest classifications can change. If in our example we used **k = 3**, there are two squares over the one triangle, so our model would classify the new data point as a blue square.



KNN: Logistic Regression Friendly Neighbor

If we make our k-value too large, our k-nearest neighbor classification will classify all new data types as the most dominant classification. In this final part to our example, **k = 7** means that all known data points are considered and our green circle would be classified as a red triangle.

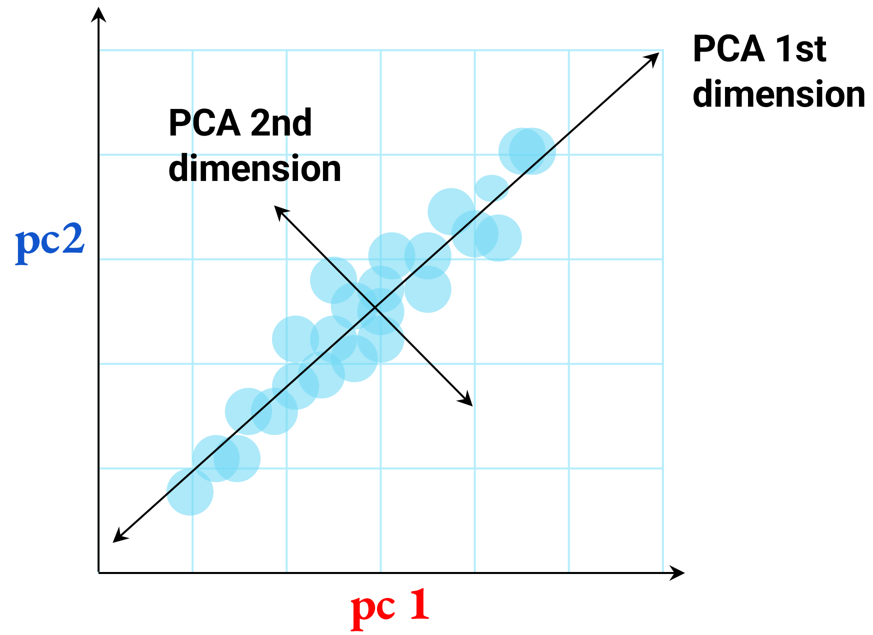


We can often enhance and optimize machine learning algorithms by applying Principal Component Analysis, or PCA.

Principal Component Analysis (PCA) is a statistical technique for streamlining the machine learning process when too many factors exist in the data.

Principal Component Analysis (PCA)

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.



Principal Component Analysis (PCA)

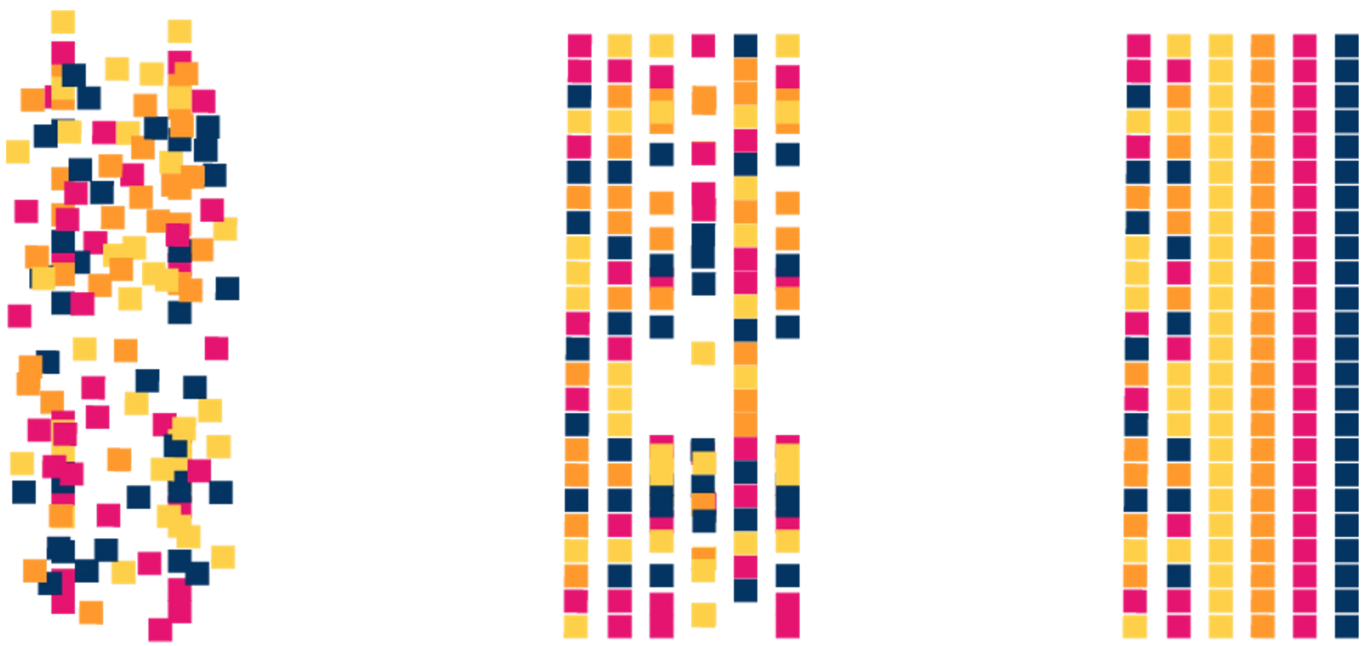
PCA is a dimensionality-reduction method that:

- Looks at all the dimensions (or data columns) in a dataset.
- Analyzes the weight of their contribution to the variance in the dataset.
- Reduces the dimensions to a smaller set that still contains as much of the information (the maximum variance) of the original dataset as possible.

PCA will NOT capture all the information from the original dataset, but it will capture as much as possible to maintain the predictive power and the meaning of the original dimensions.

Principal Component Analysis (PCA)

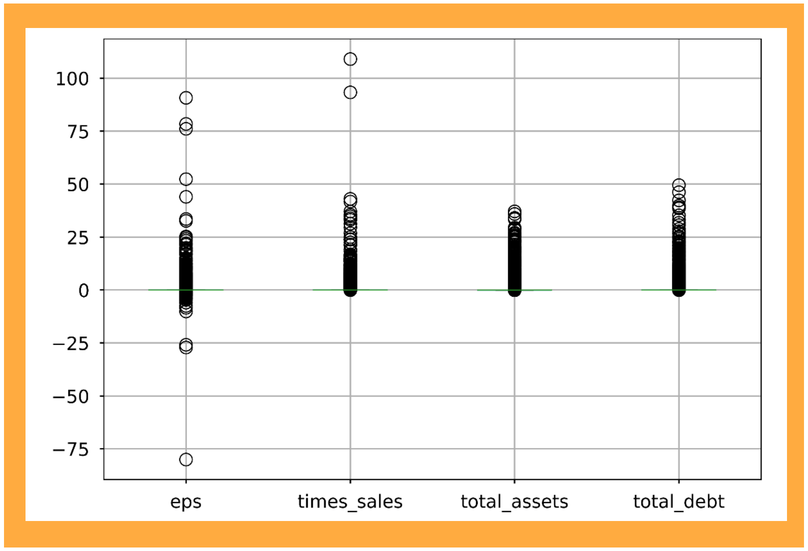
Reducing the number of factors, or **dimensional reduction**, comes at the expense of some accuracy, but the goal is to trade a little accuracy for simplicity.



Standard Scaling

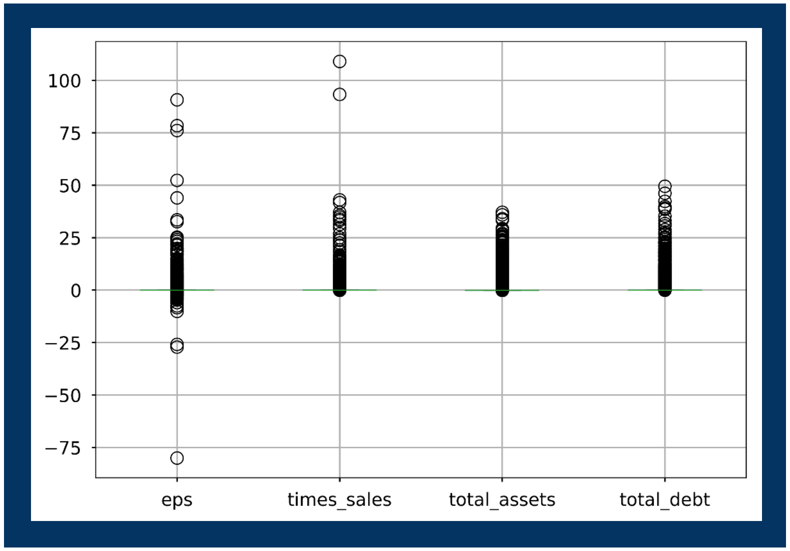
Before

Before using PCA, we'll apply standard scaling to learn how to transform the features of data.



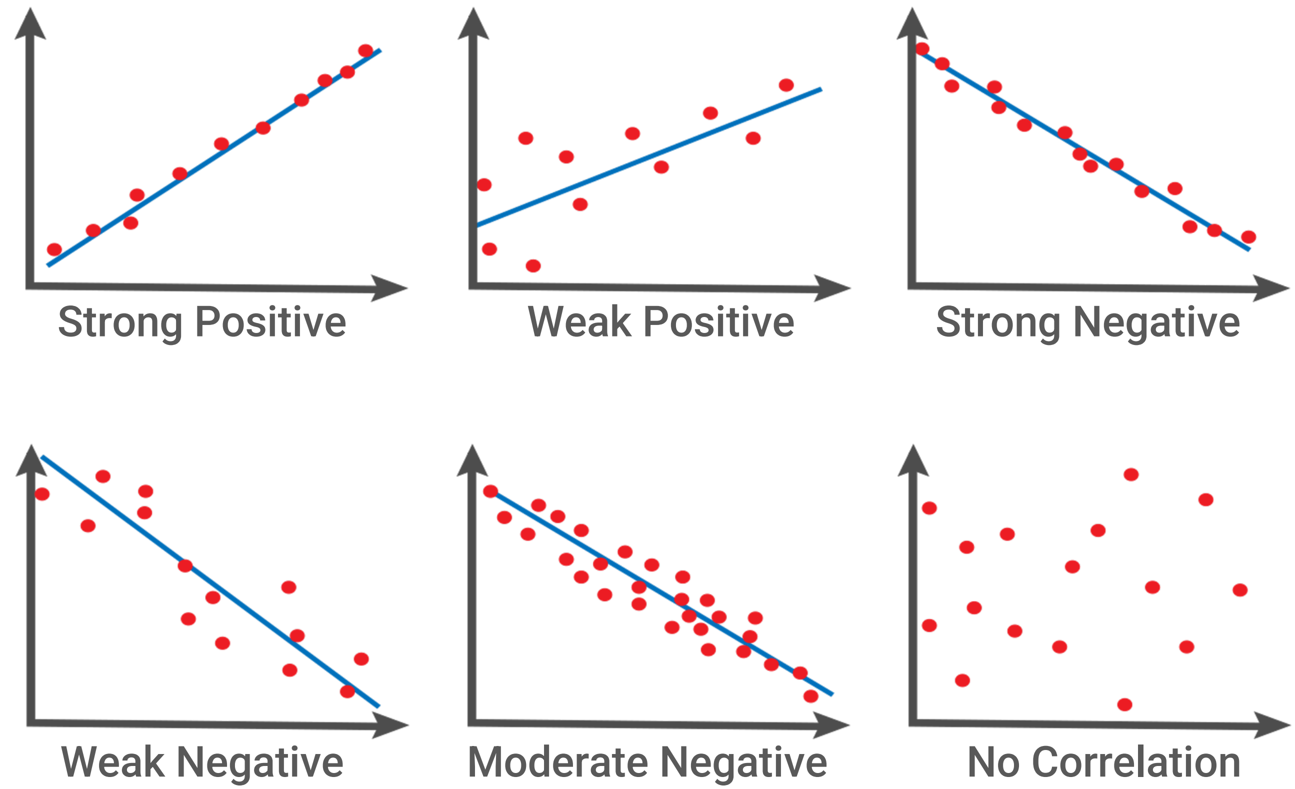
After

After scaling, we'll combine PCA with the K-means algorithm. This will give us a strategy to better handle extremely large financial datasets.



In statistics, a **correlation** defines the relationship between two or more variables.

Comparison of Correlation Relationships



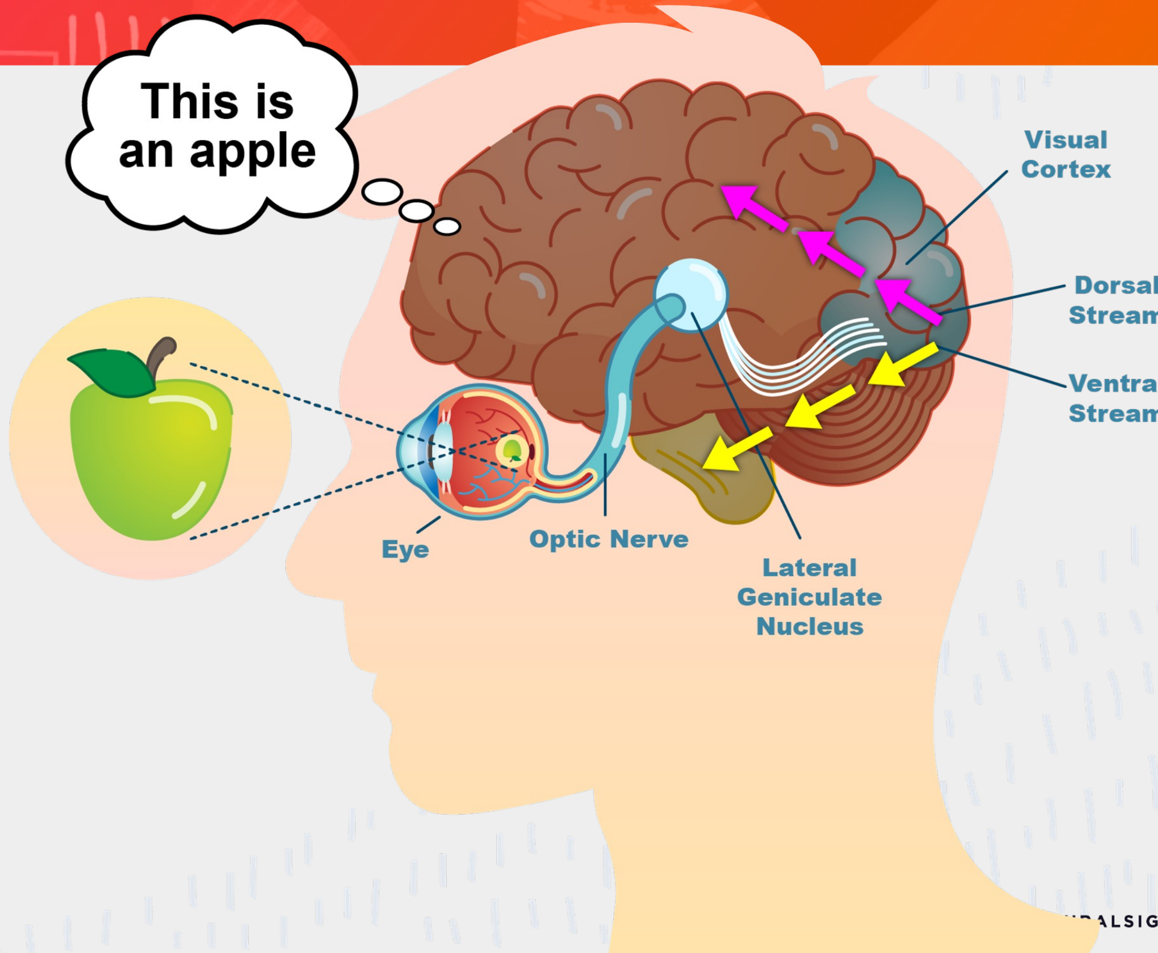
Correlations can be helpful, but they don't provide enough information to infer the relationship between two variables.

Artificial Neural Networks (ANN)

Neural Networks

How our brain works:

In order to recognize an image, our brain uses thousands
of neuron connections to find a match between the visual input and a mental representation of an object.

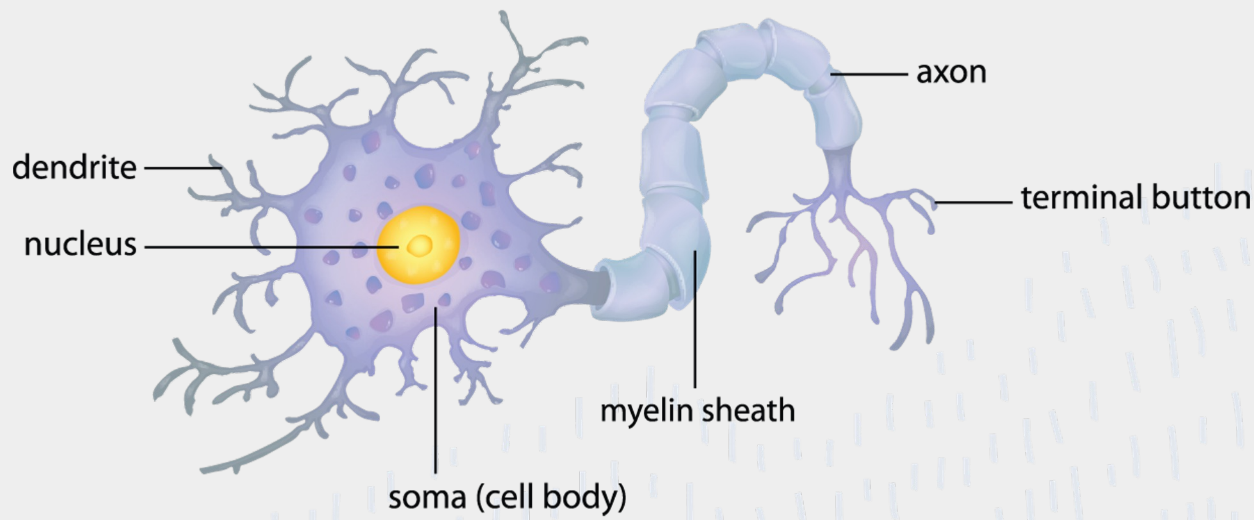


Neural Networks

The ability of the brain to process information and make predictions or interpretations is what inspired neurophysiologists and mathematicians to start the development of artificial neural networks (ANN).

In the same way that biological neurons receive input signals through the dendrites, an ANN receives input variables and processes them by using an activation function.

The output of an ANN is similar to the neuron nucleus in the brain.



History of Neural Networks

1943

Neurophysiologist **Warren McCulloch** and mathematician **Walter Pitts** wrote a paper on how neurons might work.

1949

Donald Hebb wrote *The Organization of Behavior*, which pointed out the fact that neural pathways are strengthened each time they are used.

1959

Bernard Widrow and **Marcian Hoff** of Stanford developed models called ADALINE and MADALINE.

1962

Widrow and **Hoff** developed a learning procedure that examines the value before the weight adjusts it (i.e., 0 or 1) according to the rule: Weight Change = (Pre-Weight line value).

1972

Teuvo Kohonen and **James A. Anderson** each developed a similar network independently of one another. They both used matrix mathematics to describe their ideas but did not realize that what they were doing was creating an array of analog ADALINE circuits.

History of Neural Networks

1982

John Hopfield of Caltech presented a paper to the National Academy of Sciences. His approach was to create more useful machines by using bidirectional lines. Previously, the connections between neurons was only one way.

1982

Joint US-Japan conference on **Cooperative/ Competitive Neural Networks**. Japan announced a new Fifth Generation effort on neural networks, and US papers generated worry that the US could be left behind in the field.

1986

Three independent groups of researchers, including **David Rumelhart**, a former member of Stanford's psychology department, came up with similar ideas which are now called back propagation networks.

1997

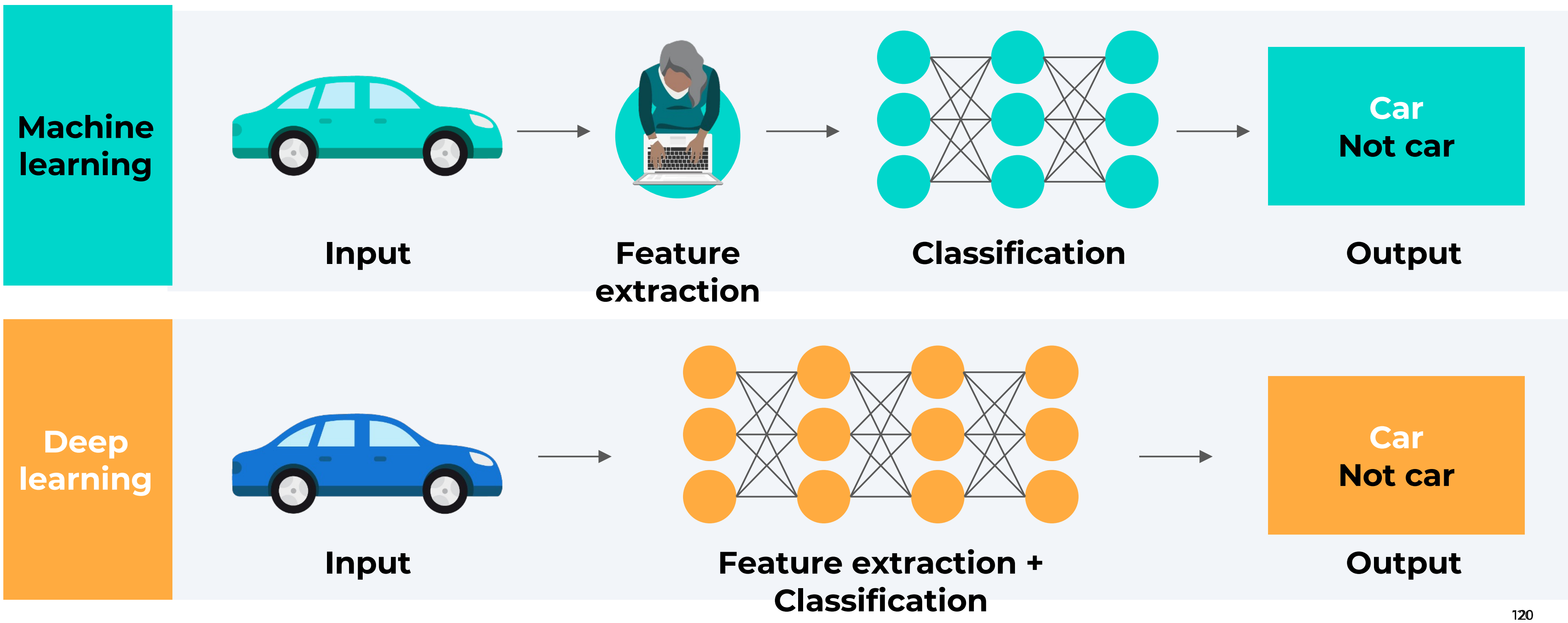
A recurrent neural network framework, LSTM was proposed by **Jürgen Schmidhuber** and **Sepp Hochreiter**.

2000s

Transformers were introduced. Followed by **GANs**, **VAEs**, and **Autoregressive** models which pushed the boundaries of **Generative AI**.

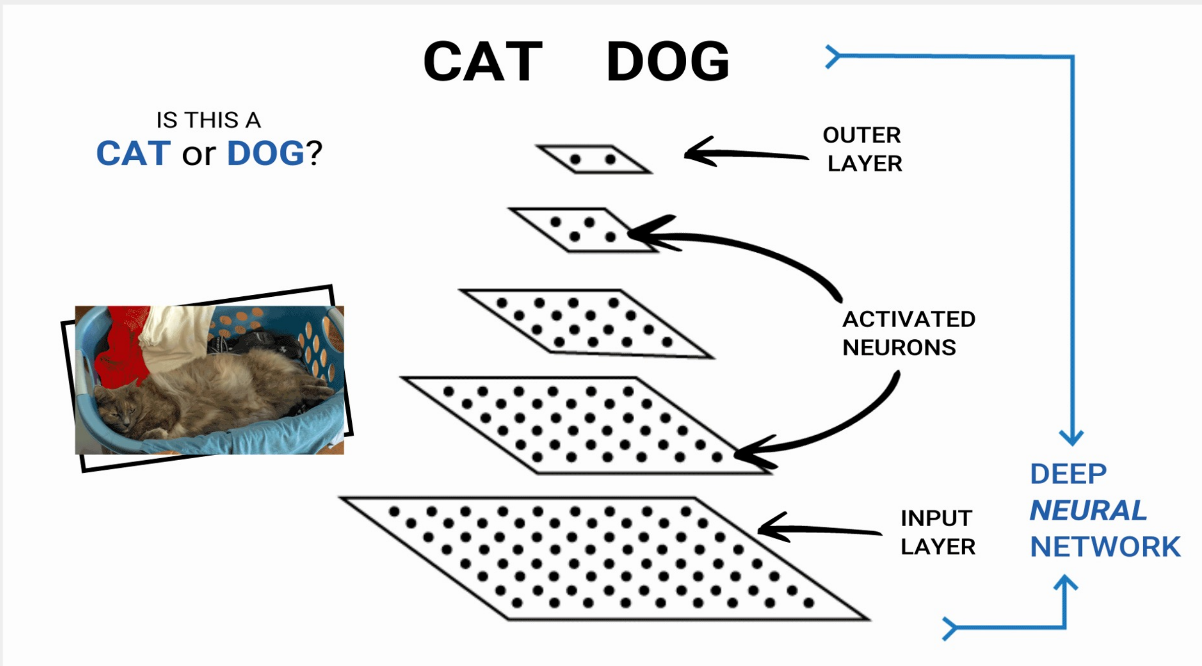
Machine Learning vs. Deep Learning

Deep neural networks are much more effective than traditional machine-learning approaches at discovering nonlinear relationships among data.



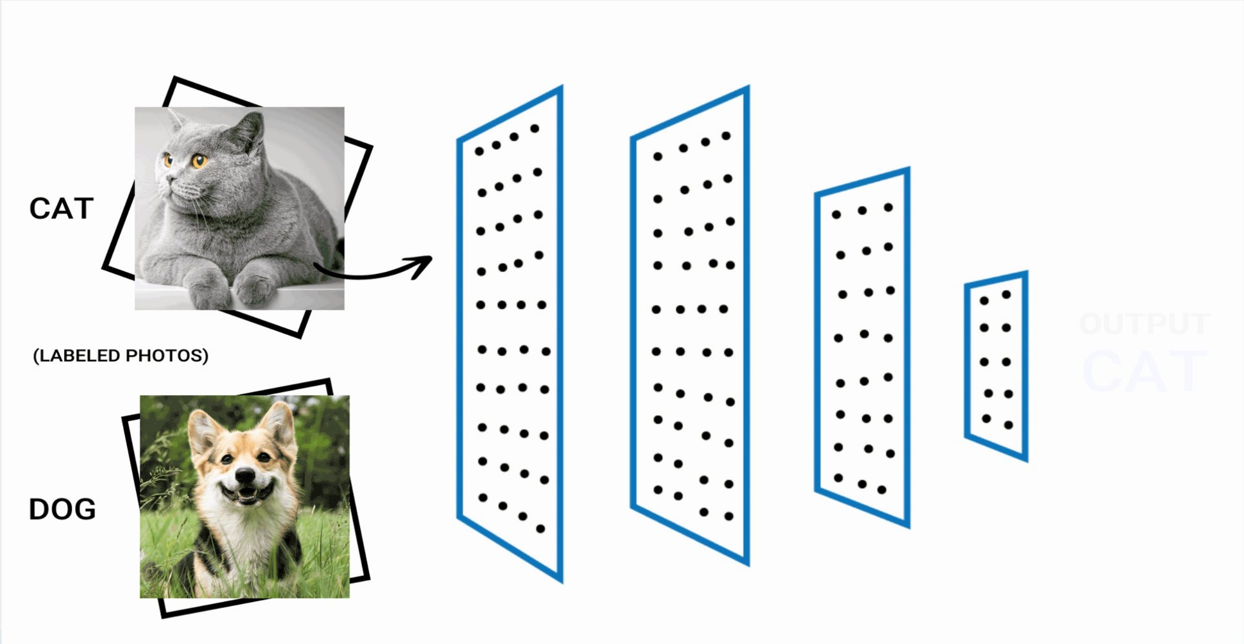
Neural Networks

Neural networks calculate the weights of various input data and pass them to the next layer of neurons. This process continues until the data reaches the output layer, which makes the final decision on the predicted category or numerical value of an instance.



Neural Networks

While definitions vary, we can consider neural networks with more than one hidden layer to be deep learning models. The decreasing cost and greater availability of computing power has increased our ability to create and use these models.



Thank you!

If you have any additional questions, please ask!If

PLURALSIGHT