Softnix

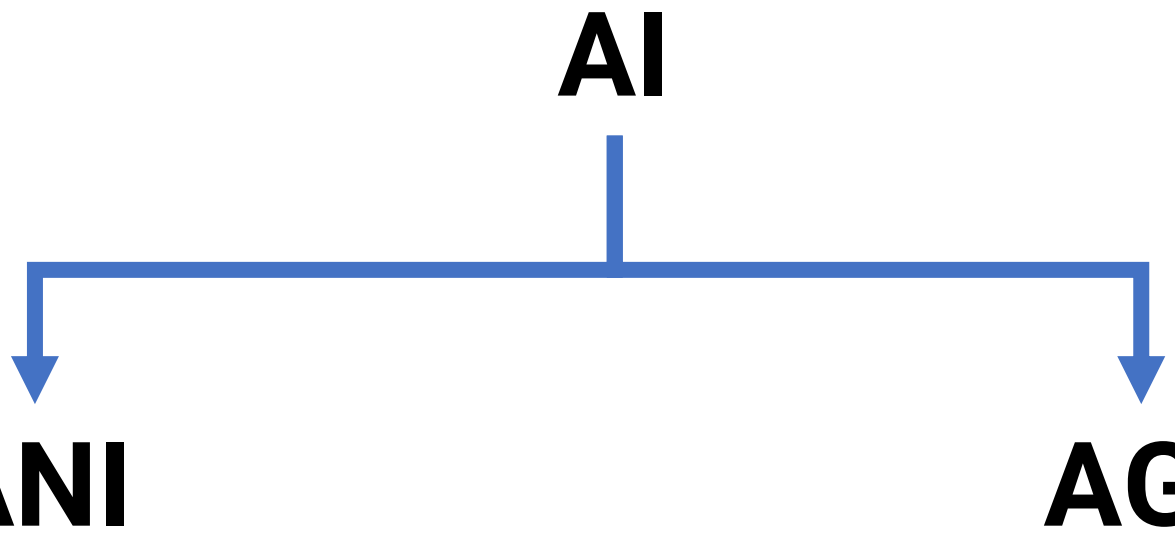
Introduction to Machine Learning

Chakrit Phain

Runchana Laohawiwat

Softnix

What is AI?



ANI

Artificial Narrow Intelligence

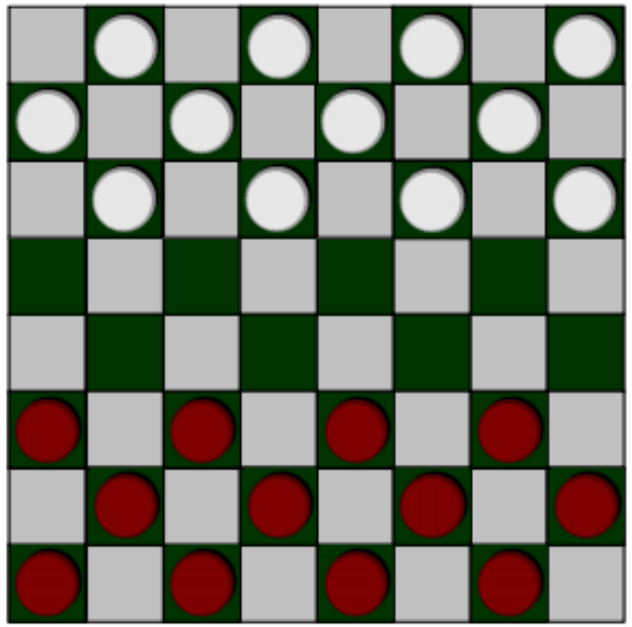
AGI

Artificial General Intelligence

https://www.coursera.org/learn/ai-for-everyone/lecture/SRwLN/week-1-introduction

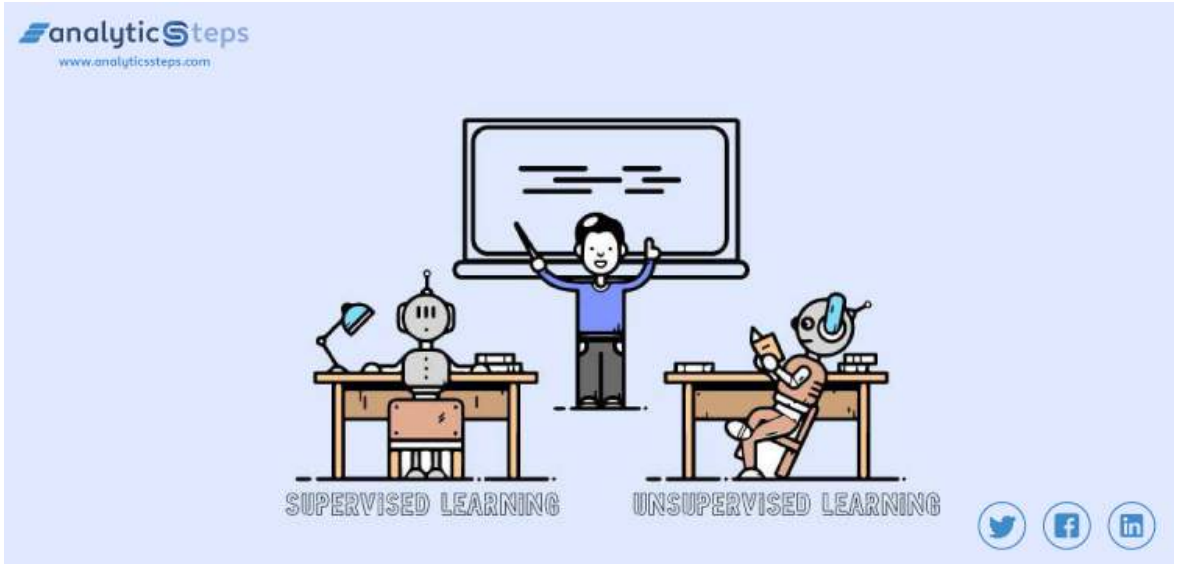
Softnix

What is Machine Learning?



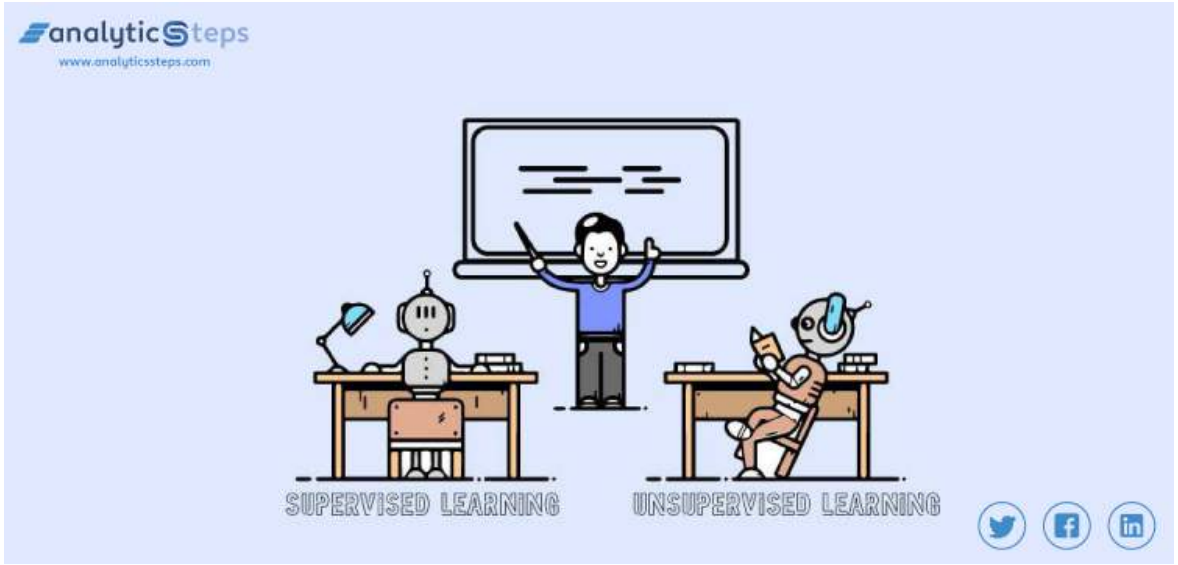
Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed.
- Arthur Samuel (1959)

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf



Machine Learning is any process by which a system improves performance from experience.
- Herbert Simon (Turing Award 1975)

https://www.seas.upenn.edu/~cis519/fall2017/lectures/01_introduction.pdf



Machine learning (**ML**) is the study of computer algorithms that improve automatically through experience.
- Tom Mitchell (1998)

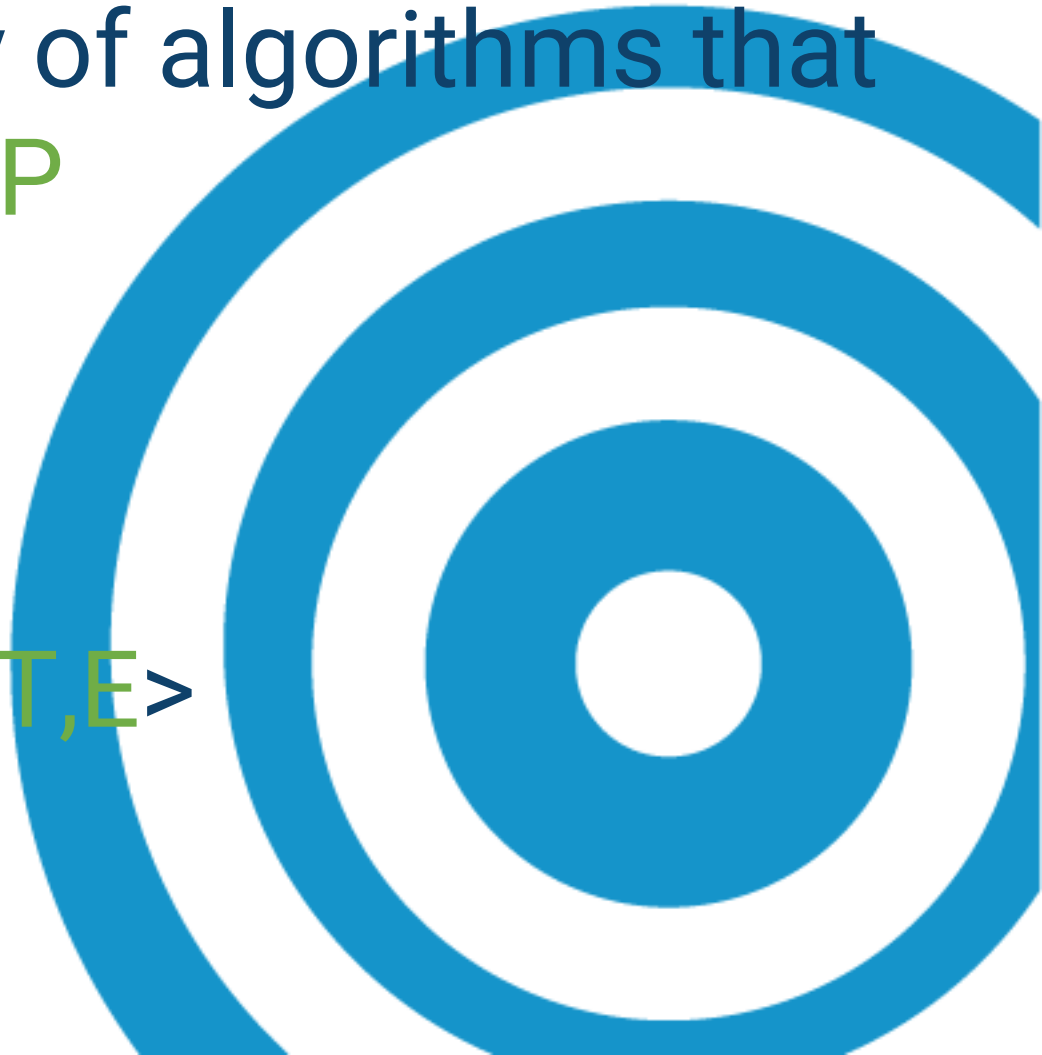
https://en.wikipedia.org/wiki/Machine_learning

Definition by Tom Mitchell (1998)

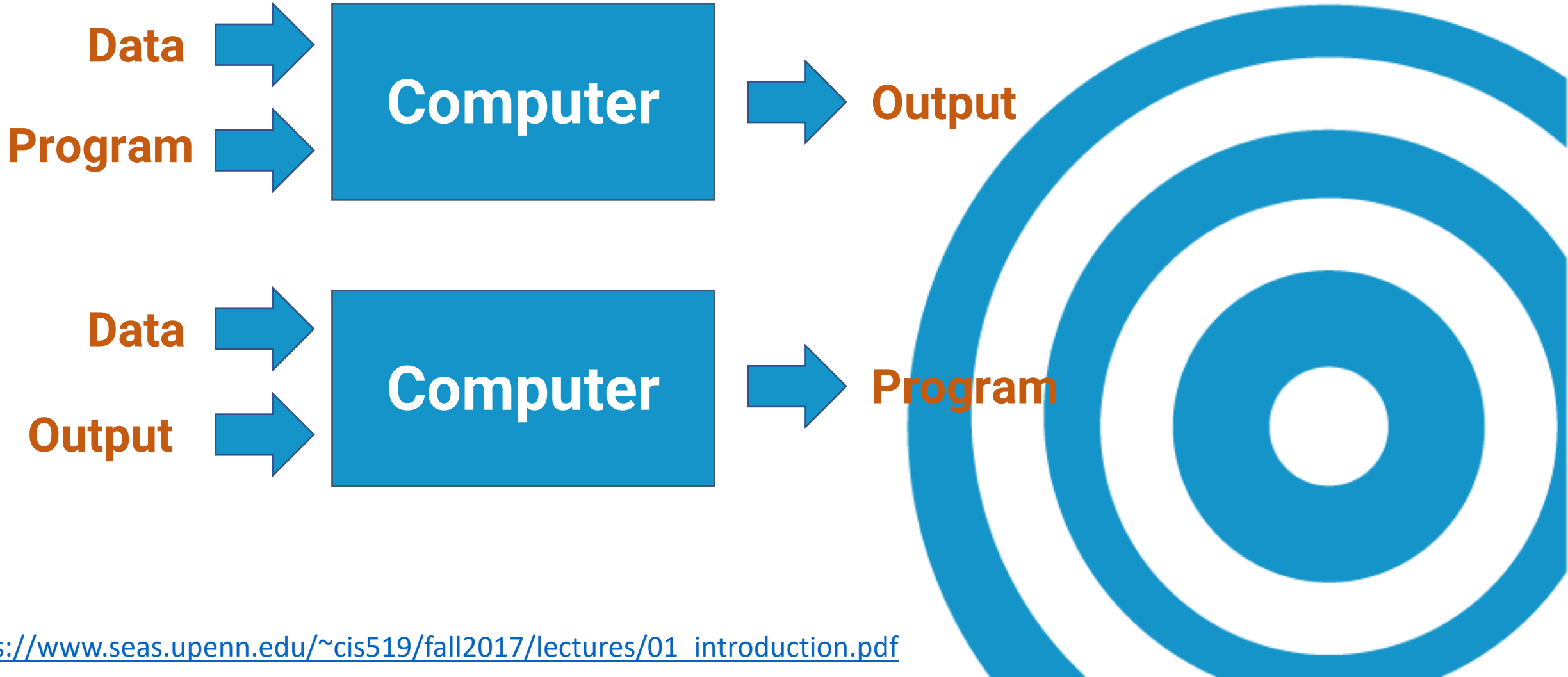
Machine Learning is the study of algorithms that

- improve their performance P
- at some task T
- with experience E

well-defined learning task: <P,T,E>

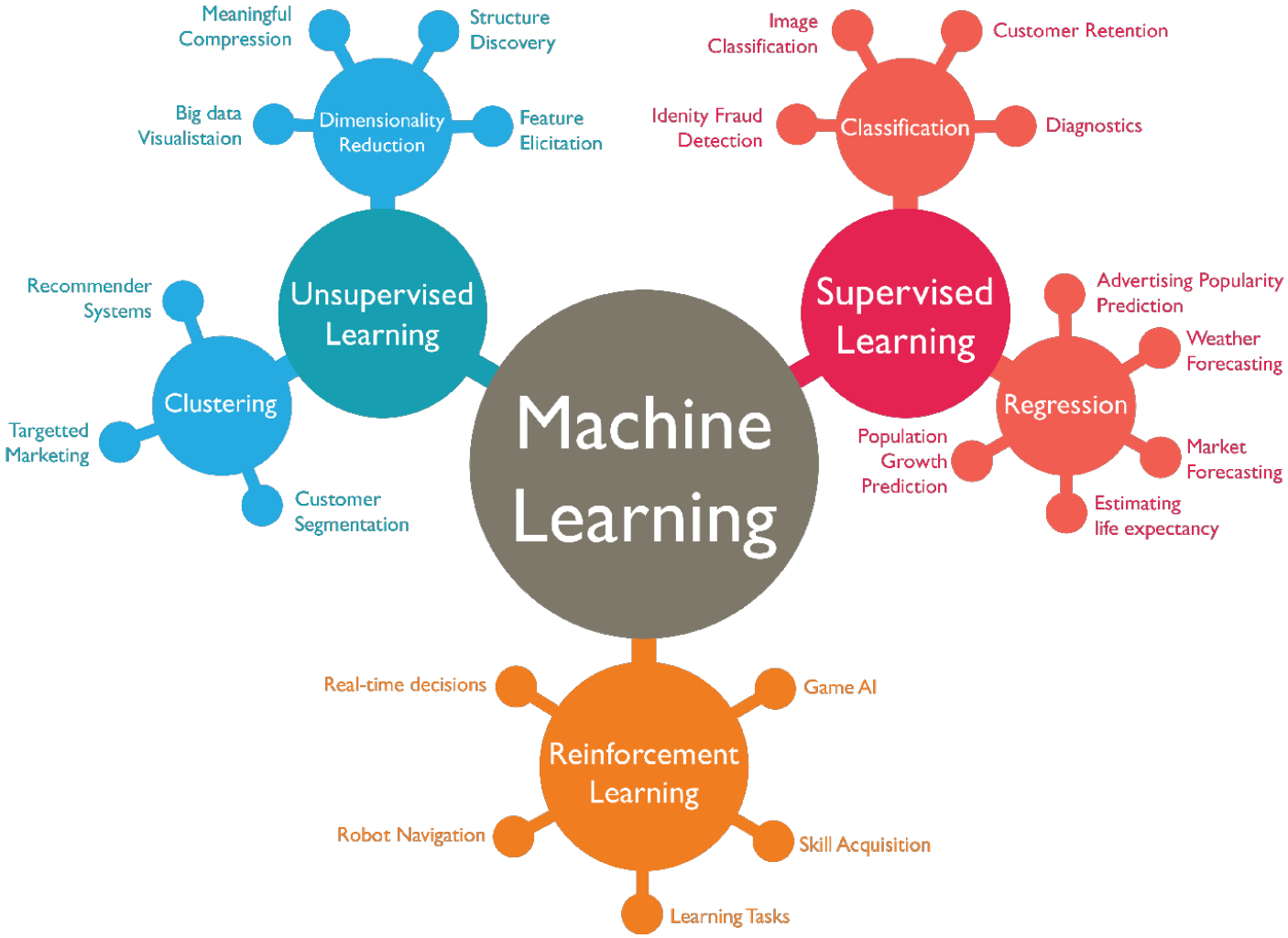


Machine Learning Definition



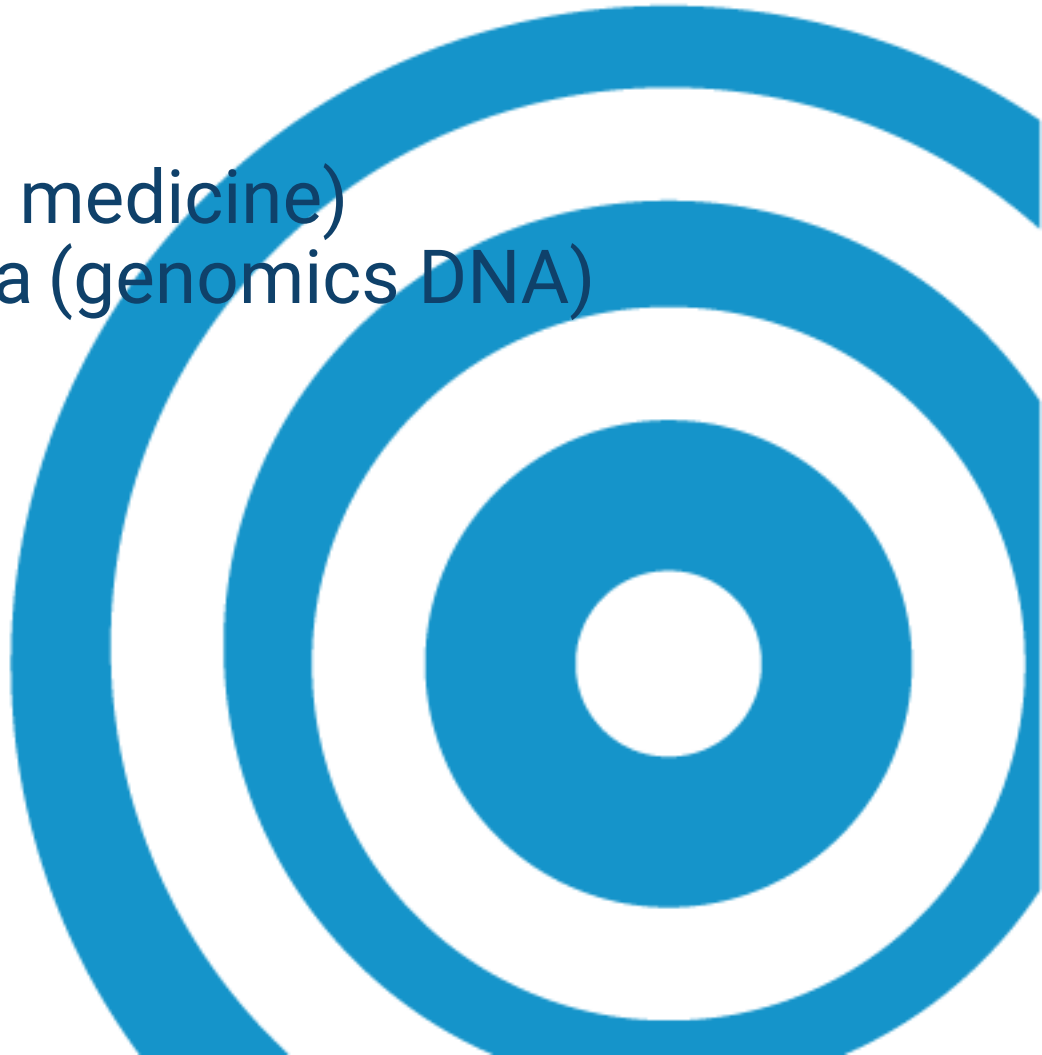
https://www.seas.upenn.edu/~cis519/fall2017/lectures/01_introduction.pdf

Machine Learning



When Do We Use Machine Learning?

- Human expertise does not exist
- Humans can’t explain their expertise
- Model must be customized (personalized medicine)
- Model are based on huge amounts of data (genomics DNA)



https://www.seas.upenn.edu/~cis519/fall2017/lectures/01_introduction.pdf

Softnix

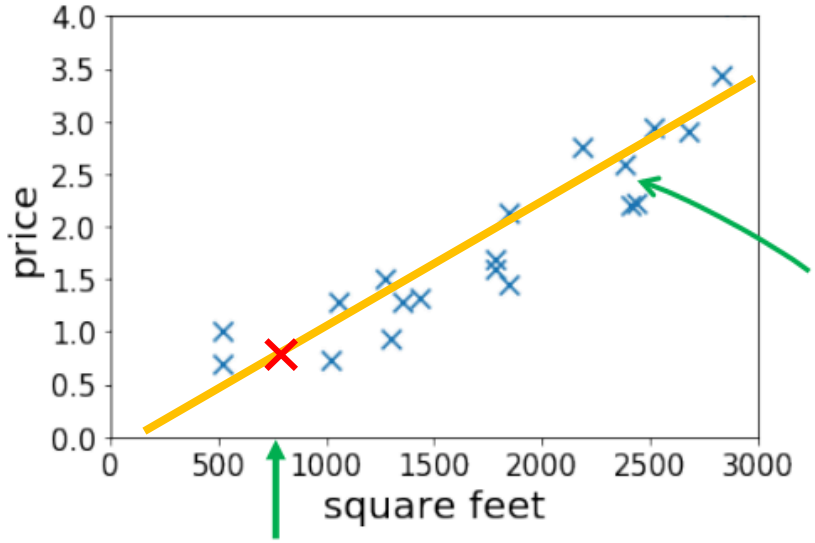
Supervised Learning

Housing Price Prediction

- Given: a dataset that contains n samples

$$(x^{(1)}, y^{(1)}), \ldots (x^{(n)}, y^{(n)})$$

- Task: if a residence has x square feet, predict its price?



15th sample
$(x^{(15)}, y^{(15)})$

$x = 800$
$y = ?$

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Housing Price Prediction

- Suppose we also know the lot size
- Task: find a function that maps

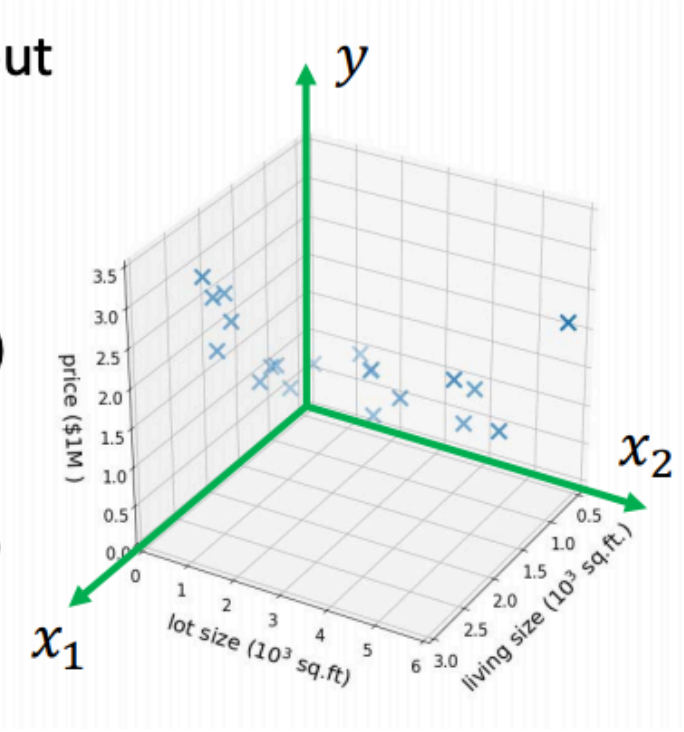
$$\underbrace{(\text{size, lot size})}_{\text{features/input}} \rightarrow \underbrace{\text{price}}_{\text{label/output}}$$

$x \in \mathbb{R}^2$ $\quad$ $y \in \mathbb{R}$

- Dataset: $\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(n)}, y^{(n)}\right)$

where $x^{(i)} = (x_1^{(i)}, x_2^{(i)})$

- "Supervision" refers to $y^{(1)}, \ldots, y^{(n)}$



http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Housing Price Prediction

- $x \in \mathbb{R}^d$ for large d
- E.g.,

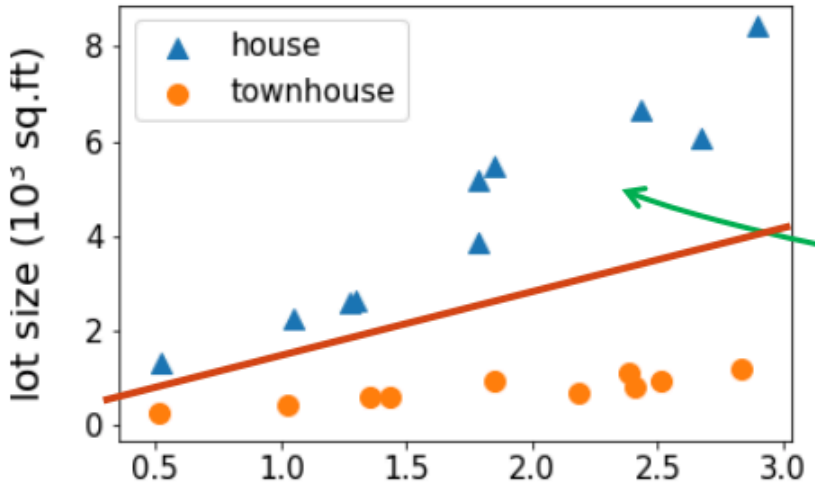
$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_d \end{bmatrix} \begin{matrix} \text{--- living size} \\ \text{--- lot size} \\ \text{--- \# floors} \\ \text{--- condition} \\ \text{--- zip code} \\ \vdots \end{matrix} \longrightarrow y \text{ --- price}$$

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Regression vs Classification

- regression: if $y \in \mathbb{R}$ is a continuous variable
  - e.g., price prediction
- classification: the label is a discrete variable
  - e.g., the task of predicting the types of residence

(size, lot size) → house or townhouse?

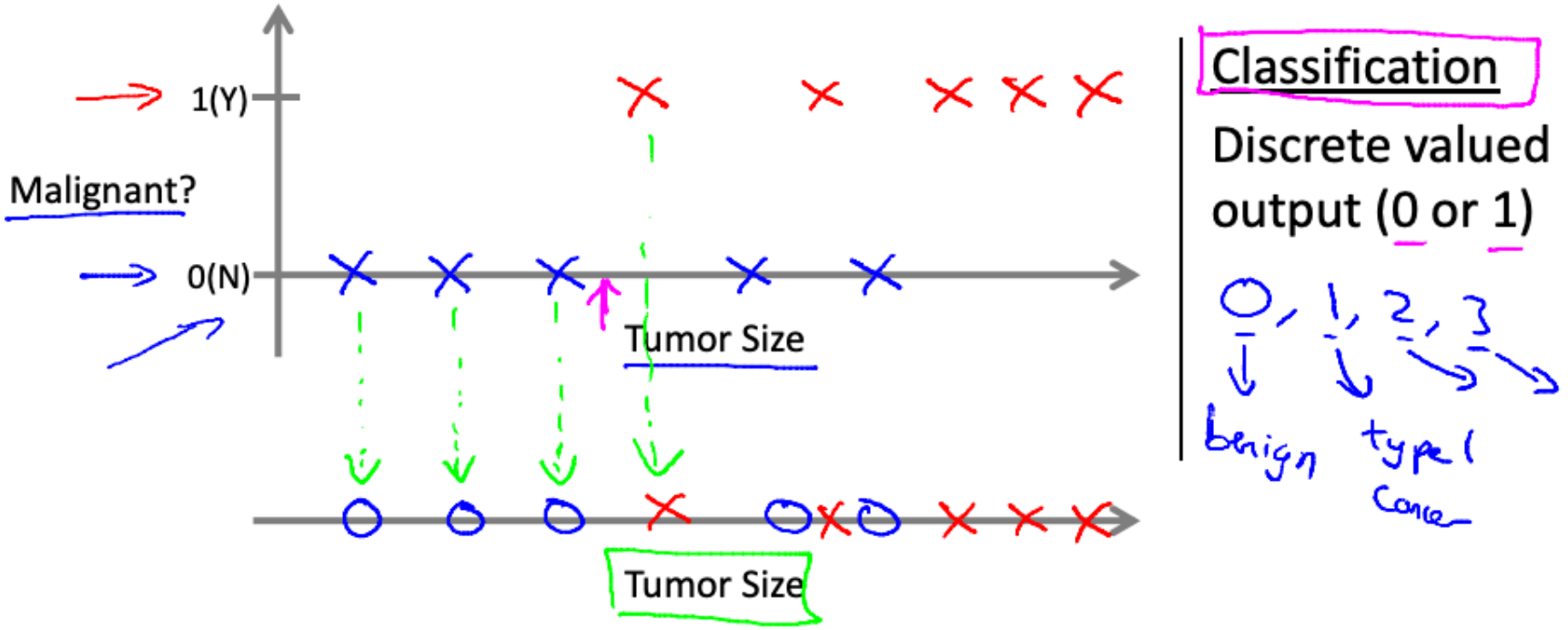


y = house or townhouse?

Lecture 3&4: classification

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Breast cancer



Classification

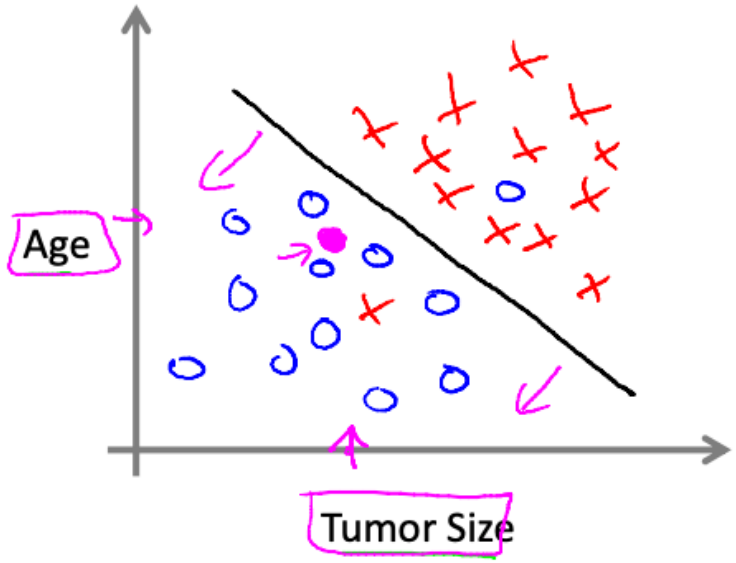
Discrete valued output (0 or 1)

0, 1, 2, 3

benign type 1 Cancer

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Breast cancer



- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

...

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

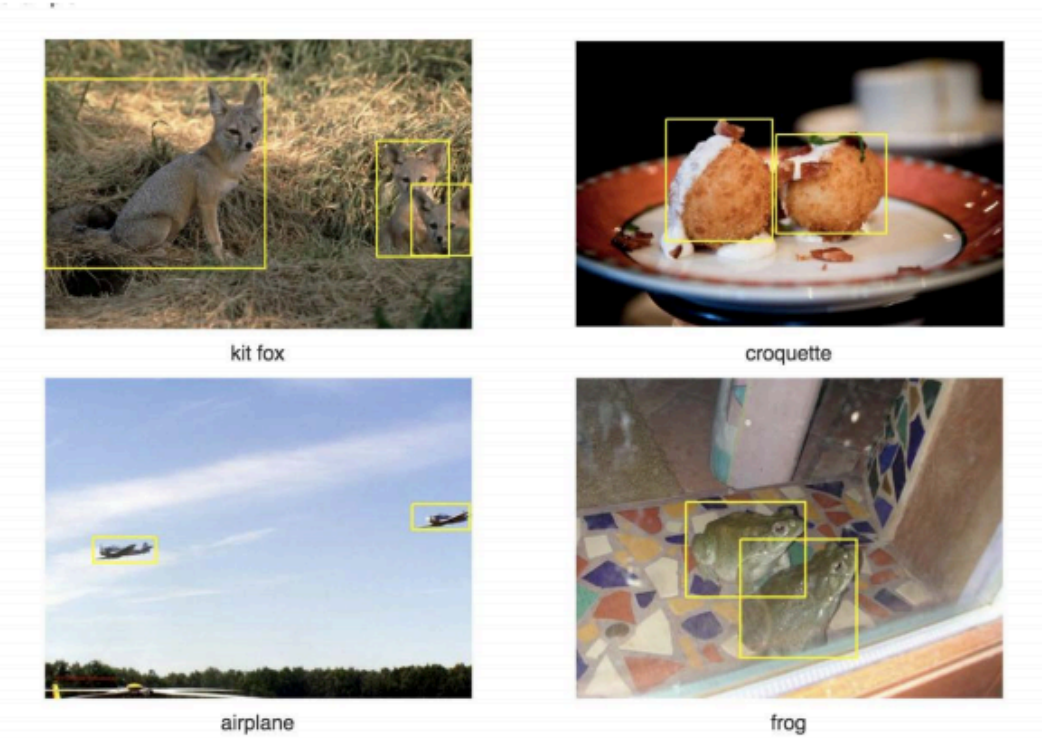
Supervised Learning in Computer Vision

- Image Classification
  - x = raw pixels of the image, y = the main object

http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Supervised Learning in Computer Vision

- Object localization and detection
  - x = raw pixels of the image, y = the bounding boxes



ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

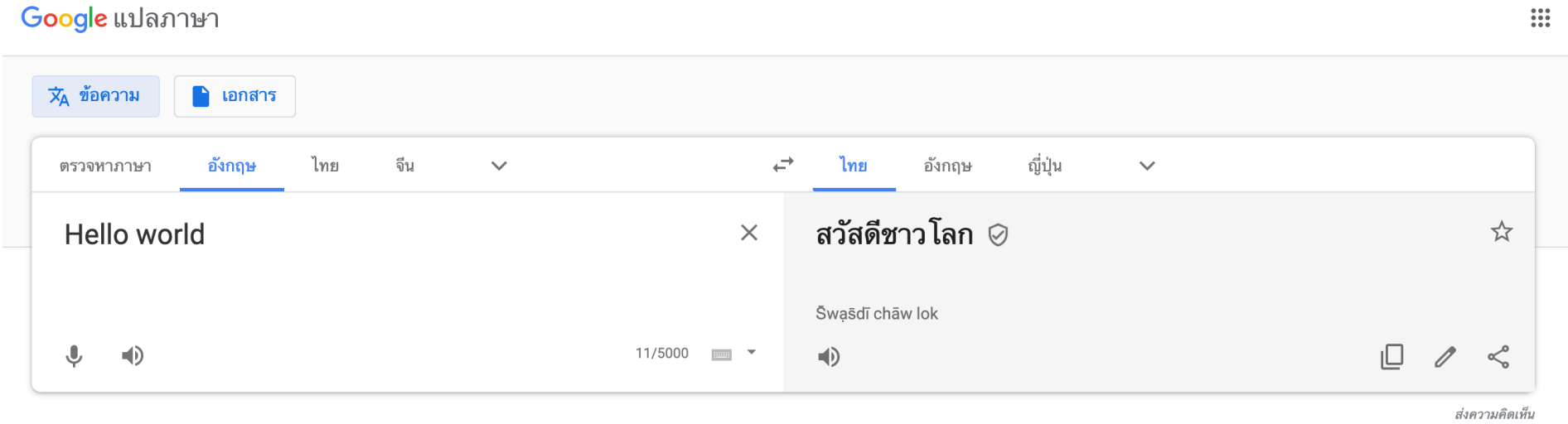
http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

Softnix

Unsupervised Learning

Supervised Learning in Natural Language Processing

Machine translation

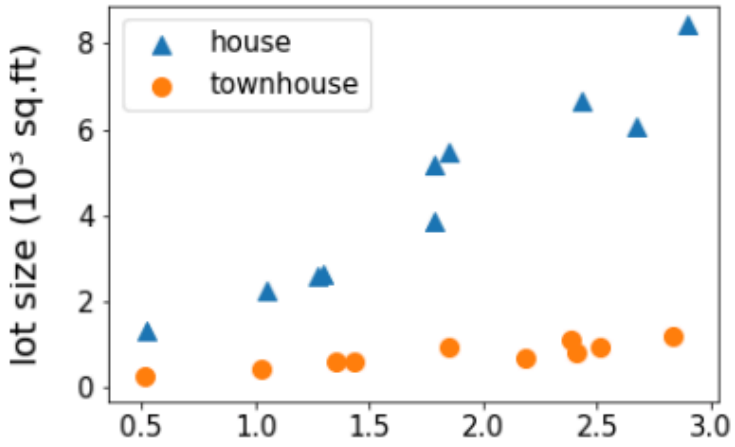


http://cs229.stanford.edu/notes2020spring/lecture1_slide.pdf

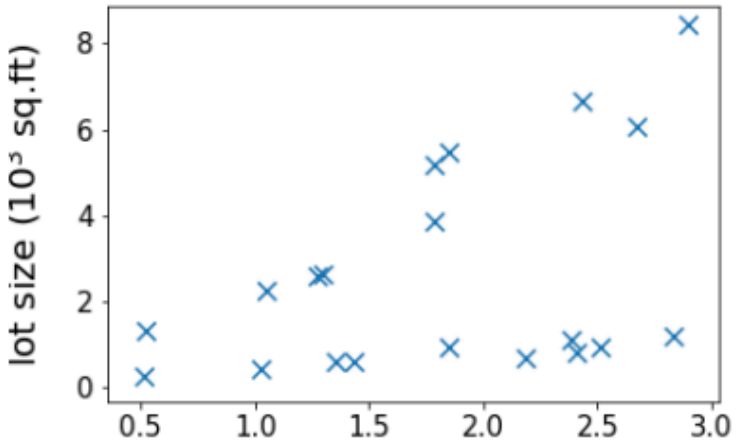
Unsupervised Learning

- Dataset contains no labels: $x^{(1)}, \ldots x^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data

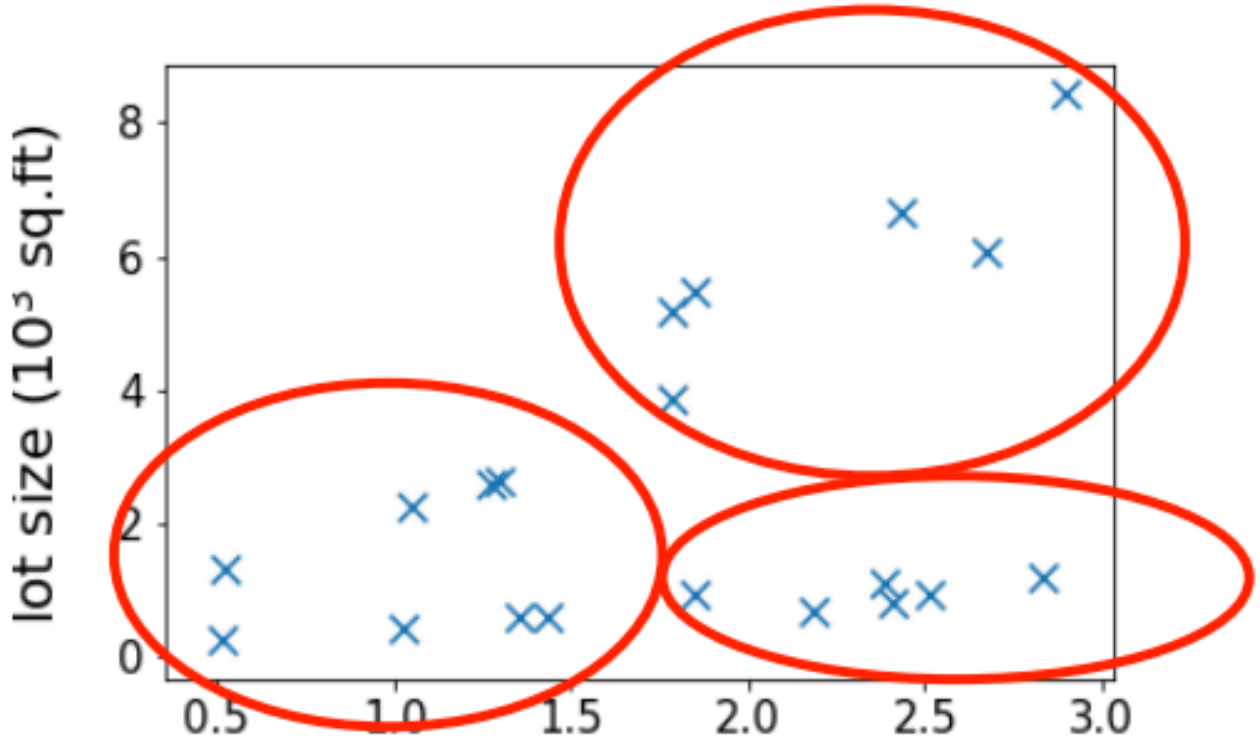
supervised



unsupervised



Clustering

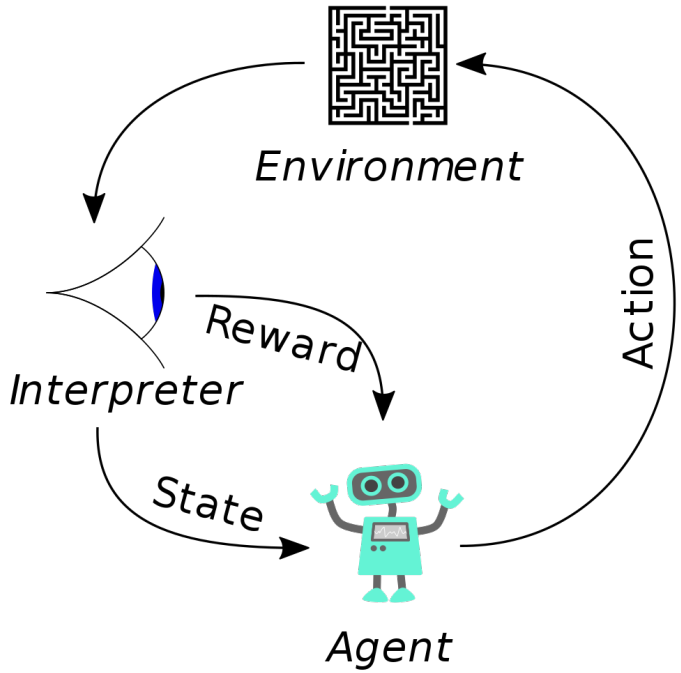


Softnix

Reinforcement Learning

Reinforcement Learning

Reinforcement learning (**RL**) is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward



https://en.wikipedia.org/wiki/Reinforcement_learning

Reinforcement Learning

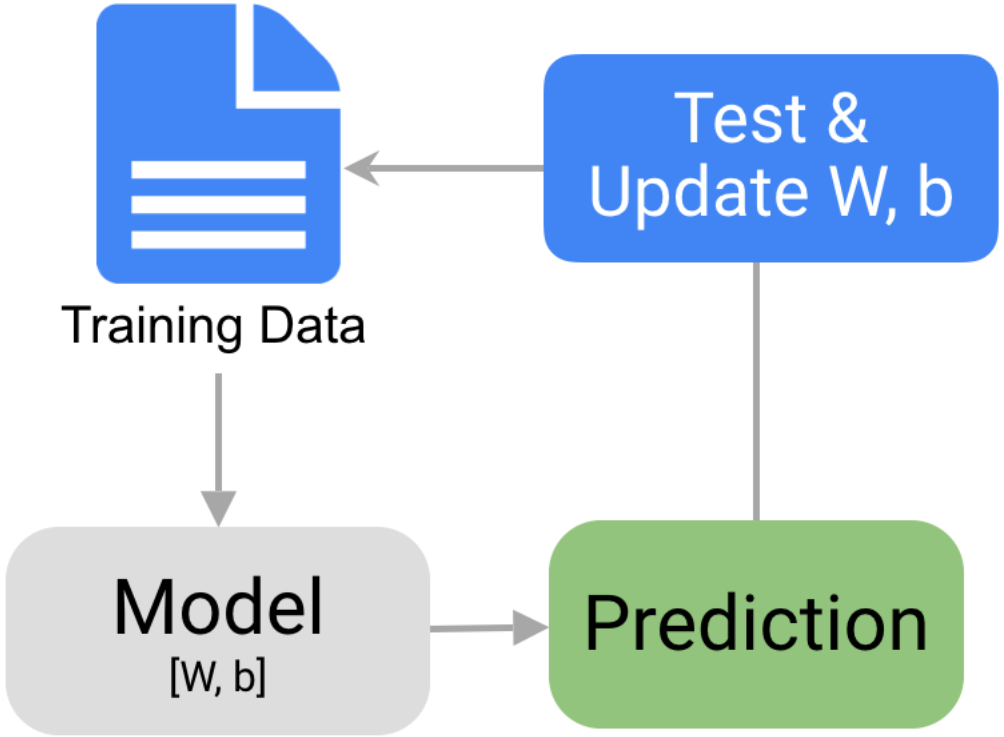
Softnix

Machine Learning Process

Machine Learning Process

https://mc.ai/7-steps-to-machine-learning-how-to-prepare-for-an-automated-future/

Machine Learning Process



https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e

Softnix

Step 2 Data preparation

Explore Data

EDA Methods

Descriptive statistics

Central tendency of a distribution

Mean	Sum of all values / Total number of values
Median	Middle value(when data are arranged in order
Mode	Most common value

Measure of Variation

Variance	**how far a set of numbers are spread out from mean**
Interquartile range	divides a data set into quartiles.
Standard deviation	dispersion of a set of data from mean

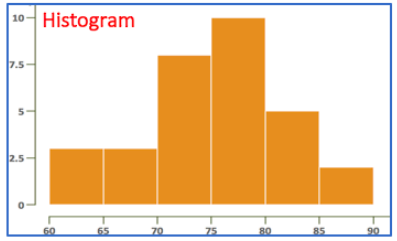
Skewness & Kurtosis

Skewness	**Measure of symmetry**
Kurtosis	**Kurtosis is a measure of "peakedness" relative to a Gaussian shape**

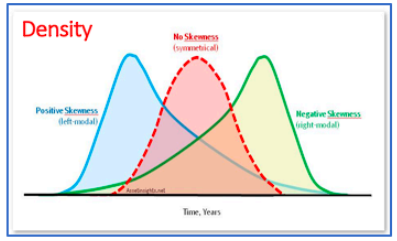
Visualizations

1-dimension

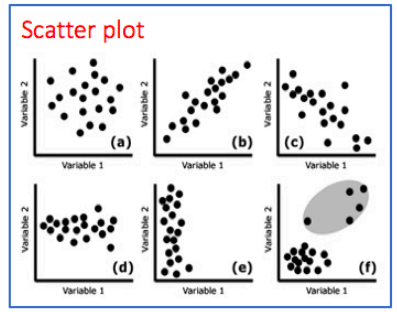
Few data points



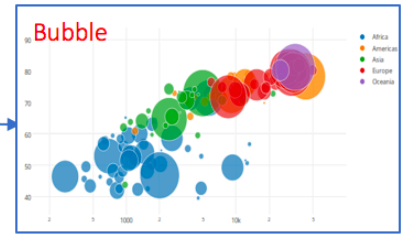
Many data points



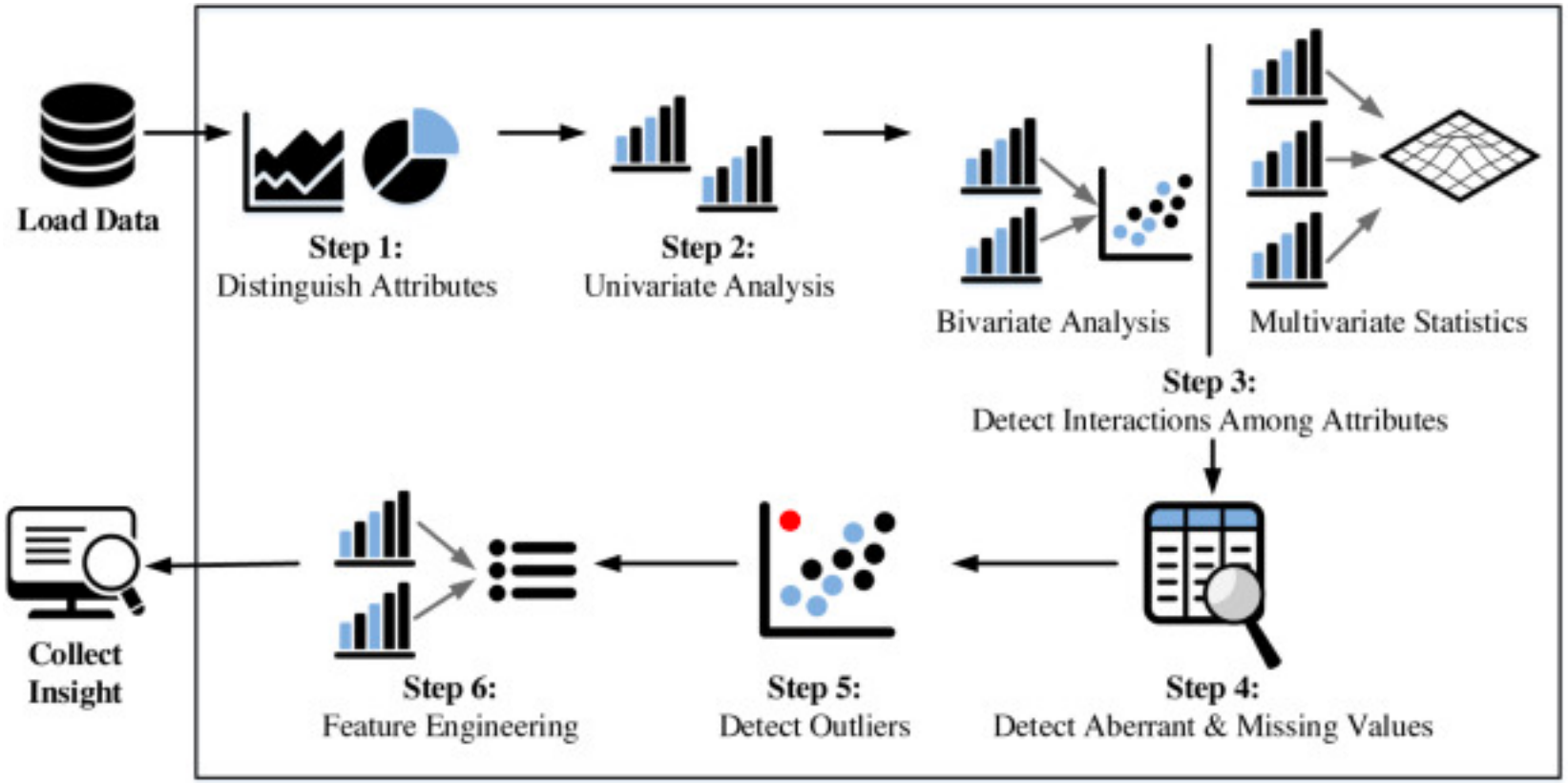
2-dimension



3-dimension

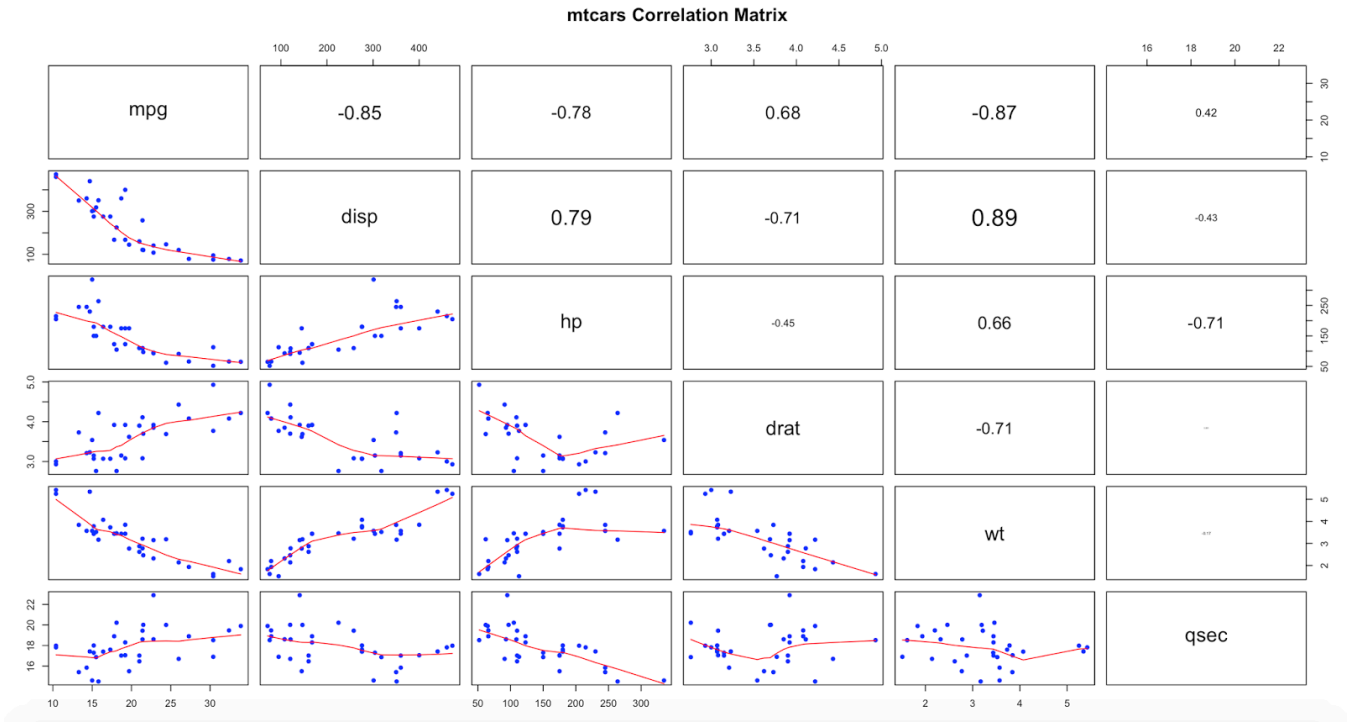


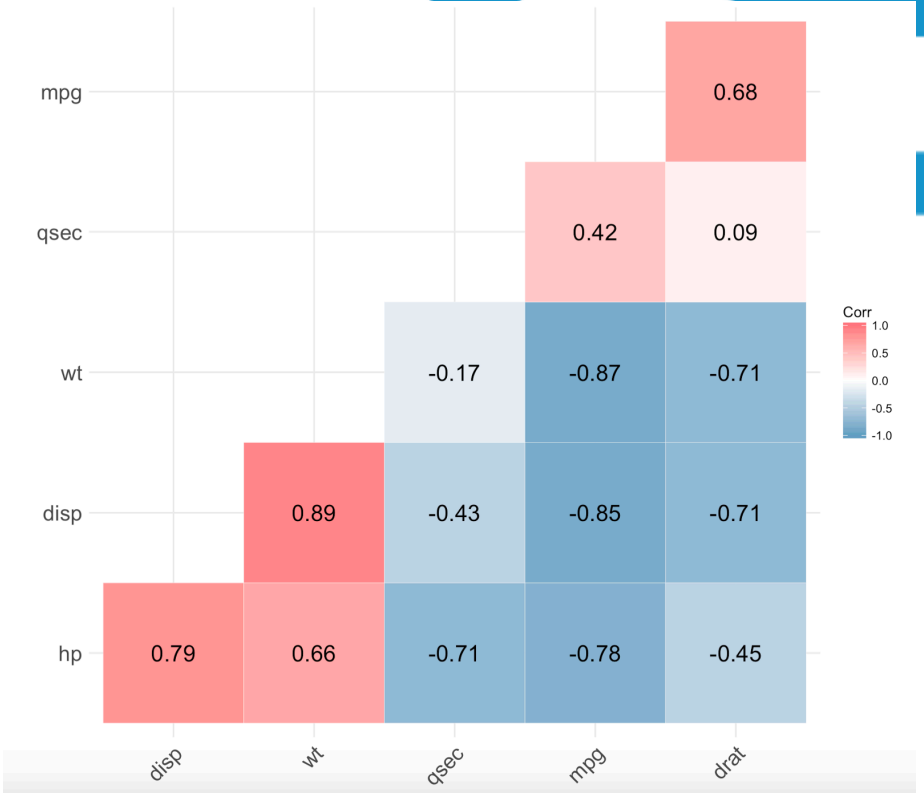
Explore Data



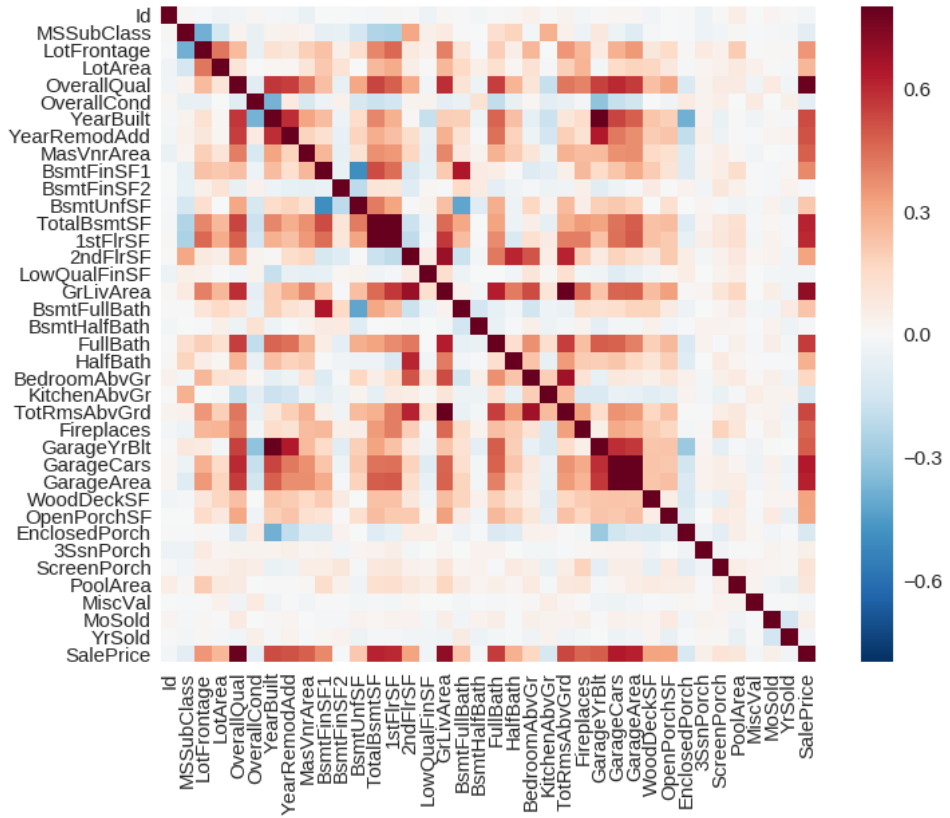
https://www.sciencedirect.com/science/article/pii/S2468502X18300561

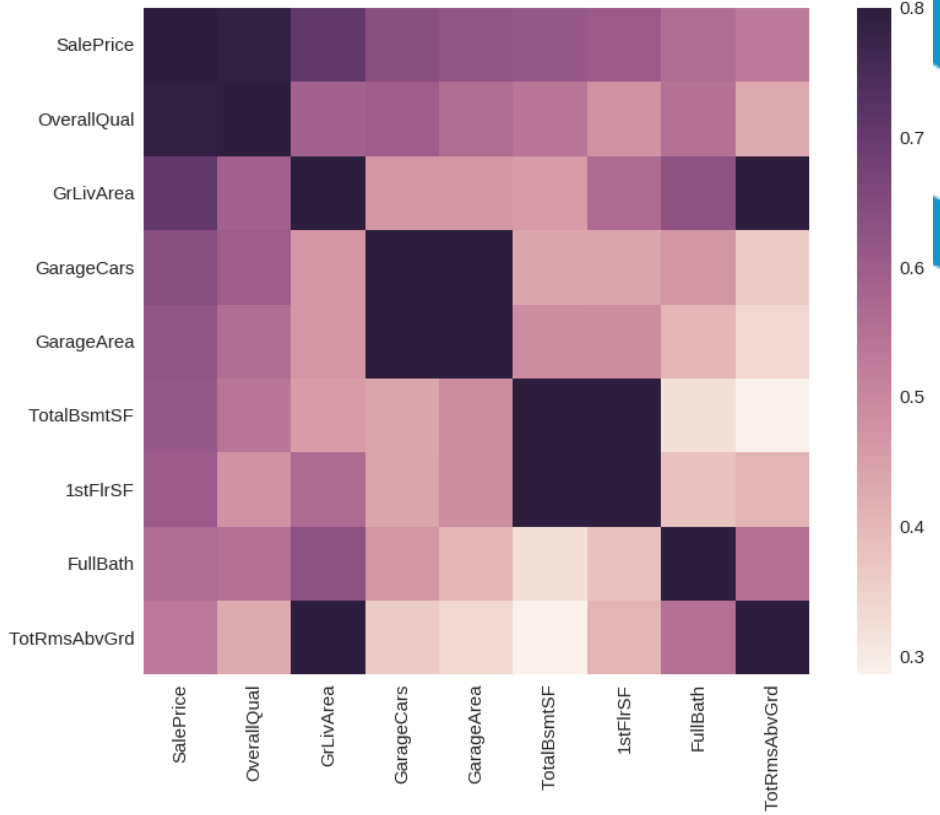
Feature Engineering





Exploratory Data Analysis Example

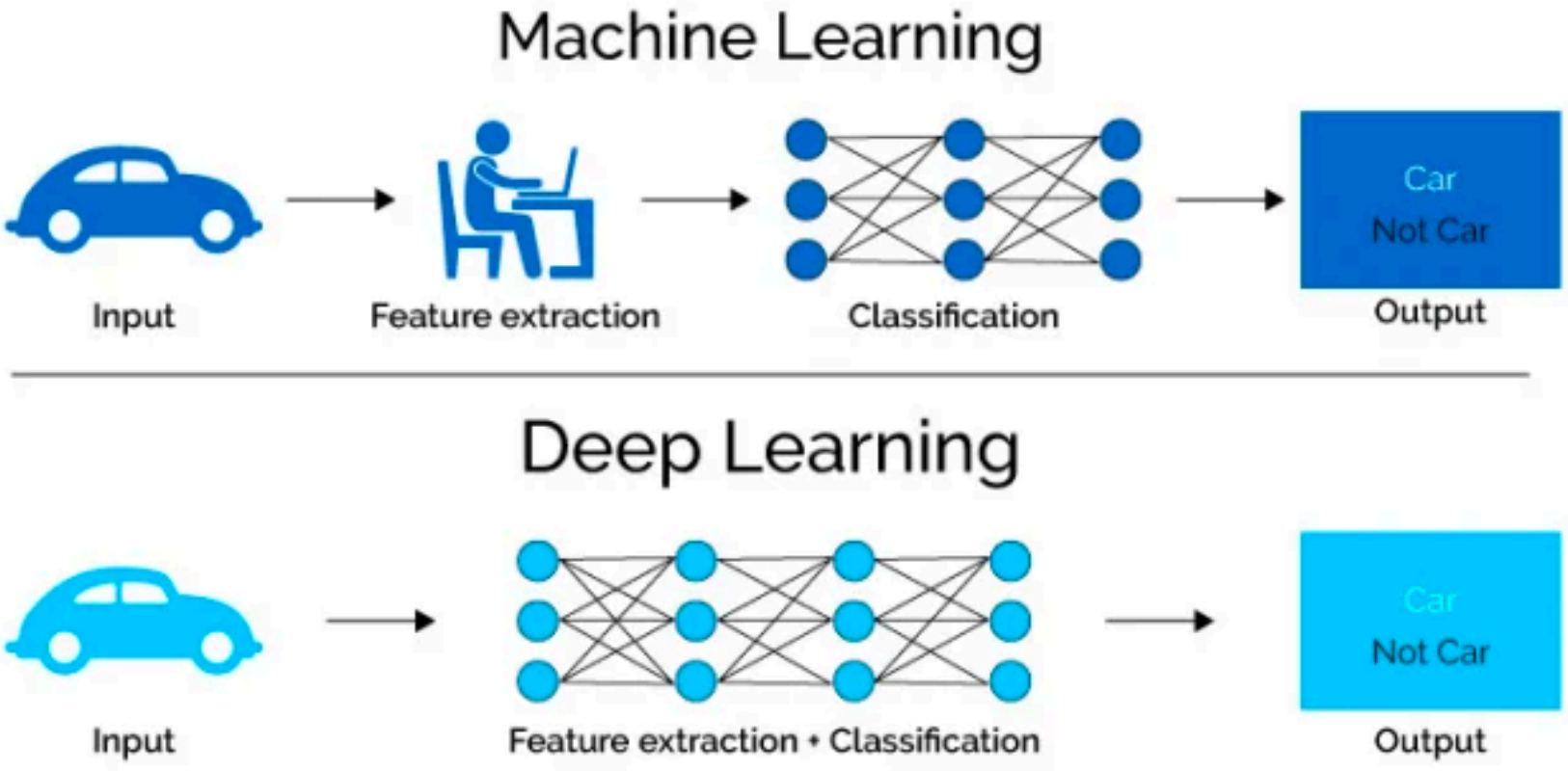




https://towardsdatascience.com/predicting-house-prices-with-linear-regression-machine-learning-from-scratch-part-ii-47a0238aeac1

https://colab.research.google.com/drive/1YEgrxbMQCONVeKDfXg49YUyD3lpwlGxu#scrollTo=DWQVJtT2uBSG

Feature Extraction



Softnix

Step 3 Choosing Model

MachineLearning Overview

MACHINE LEARNING IN EMOJI

BecomingHuman.AI

SUPERVISED — human builds model based on input / output

UNSUPERVISED — human input, machine output human utilizes if satisfactory

REINFORCEMENT — human input, machine output human reward/punish, cycle continues

BASIC REGRESSION

LINEAR
linear_model.LinearRegression()
Lots of numerical data

LOGISTIC
linear_model.LogisticRegression()
Target variable is categorical

CLUSTER ANALYSIS

K-MEANS
cluster.KMeans()
Similar datum into groups based on centroids

ANOMALY DETECTION
covariance.EllipticalEnvelope()
Finding outliers through grouping

CLASSIFICATION

NEURAL NET
neural_network.MLPClassifier()
Complex relationships. Prone to overfitting
Basically magic.

K-NN
neighbors.KNeighborsClassifier()
Group membership based on proximity

DECISION TREE
tree.DecisionTreeClassifier()
If/then/else. Non-contiguous data.
Can also be regression.

RANDOM FOREST
ensemble.RandomForestClassifier()
Find best split randomly
Can also be regression

SVM
svm.SVC() svm.LinearSVC()
Maximum margin classifier. Fundamental Data Science algorithm

NAIVE BAYES
GaussianNB() MultinomialNB() BernoulliNB()
Updating knowledge step by step with new info

FEATURE REDUCTION

T-DISTRIB STOCHASTIC NEIB EMBEDDING
manifold.TSNE()
Visual high dimensional data. Convert similarity to joint probabilities

PRINCIPLE COMPONENT ANALYSIS
decomposition.PCA()
Distill feature space into components that describe greatest variance

CANONICAL CORRELATION ANALYSIS
decomposition.CCA()
Making sense of cross-correlation matrices

LINEAR DISCRIMINANT ANALYSIS
lda.LDA()
Linear combination of features that separates classes

OTHER IMPORTANT CONCEPTS

BIAS VARIANCE TRADEOFF

UNDERFITTING / OVERFITTING

INERTIA

ACCURACY FUNCTION
(TP+TN) / (P+N)

PRECISION FUNCTION
manifold.TSNE()

SPECIFICITY FUNCTION
TN / (FP+TN)

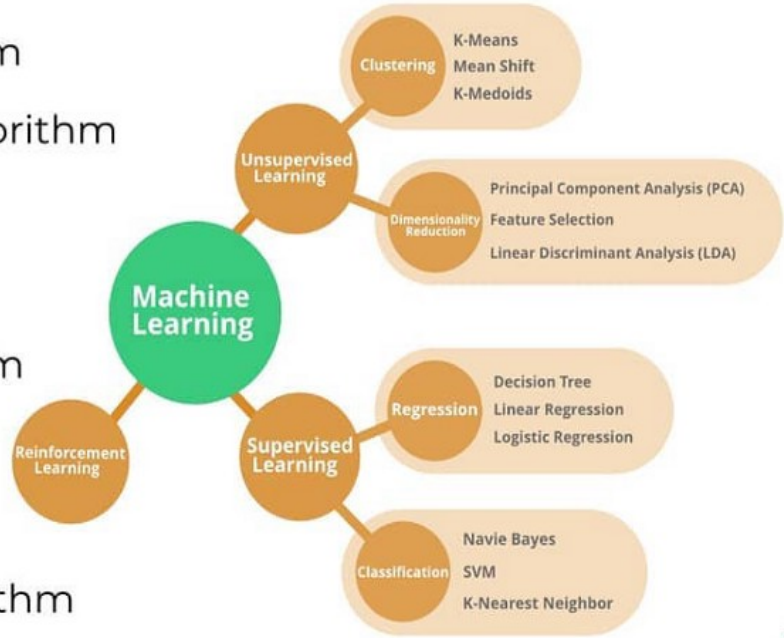
SENSITIVITY FUNCTION
TP / (TP+FN)

Originally Created by Emily Barry. **See original here.**

Softnix

Top 10 Algorithms every Machine Learning Engineer should know

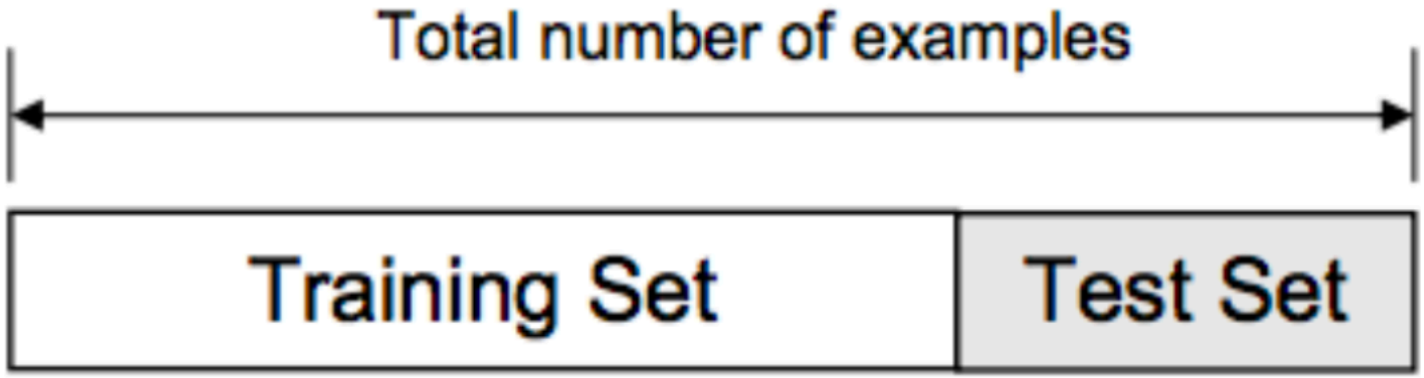
1. Naïve Bayes Classifier Algorithm
2. K Means Clustering Algorithm
3. Support Vector Machine Algorithm
4. Apriori Algorithm
5. Linear Regression Algorithm
6. Logistic Regression Algorithm
7. Decision Trees Algorithm
8. Random Forests Algorithm
9. K Nearest Neighbours Algorithm
10. Artificial Neural Networks Algorithm



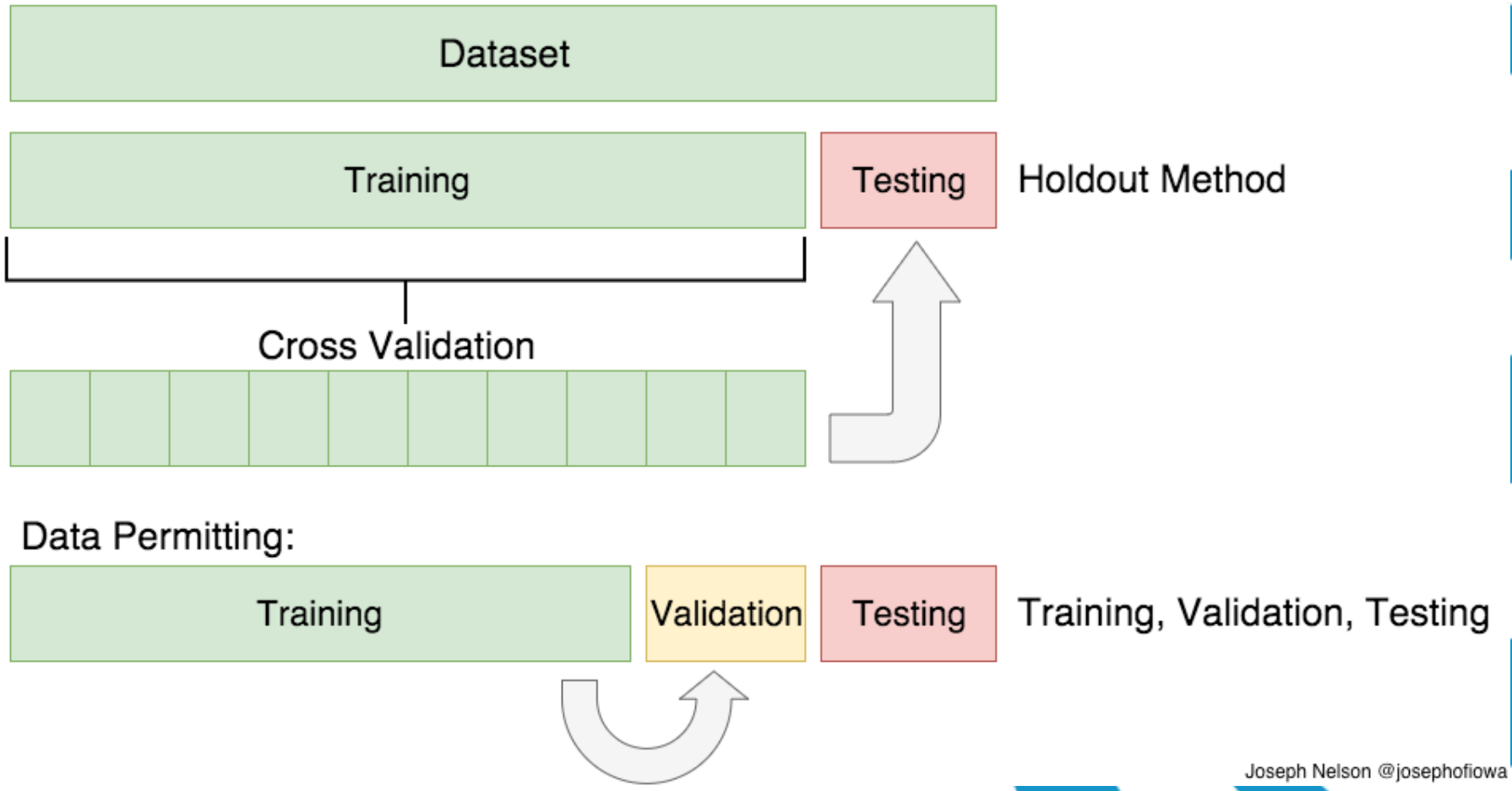
https://mc.ai/top-10-best-machine-learning-algorithms-2/

Step 4 Training

Train/Test Split

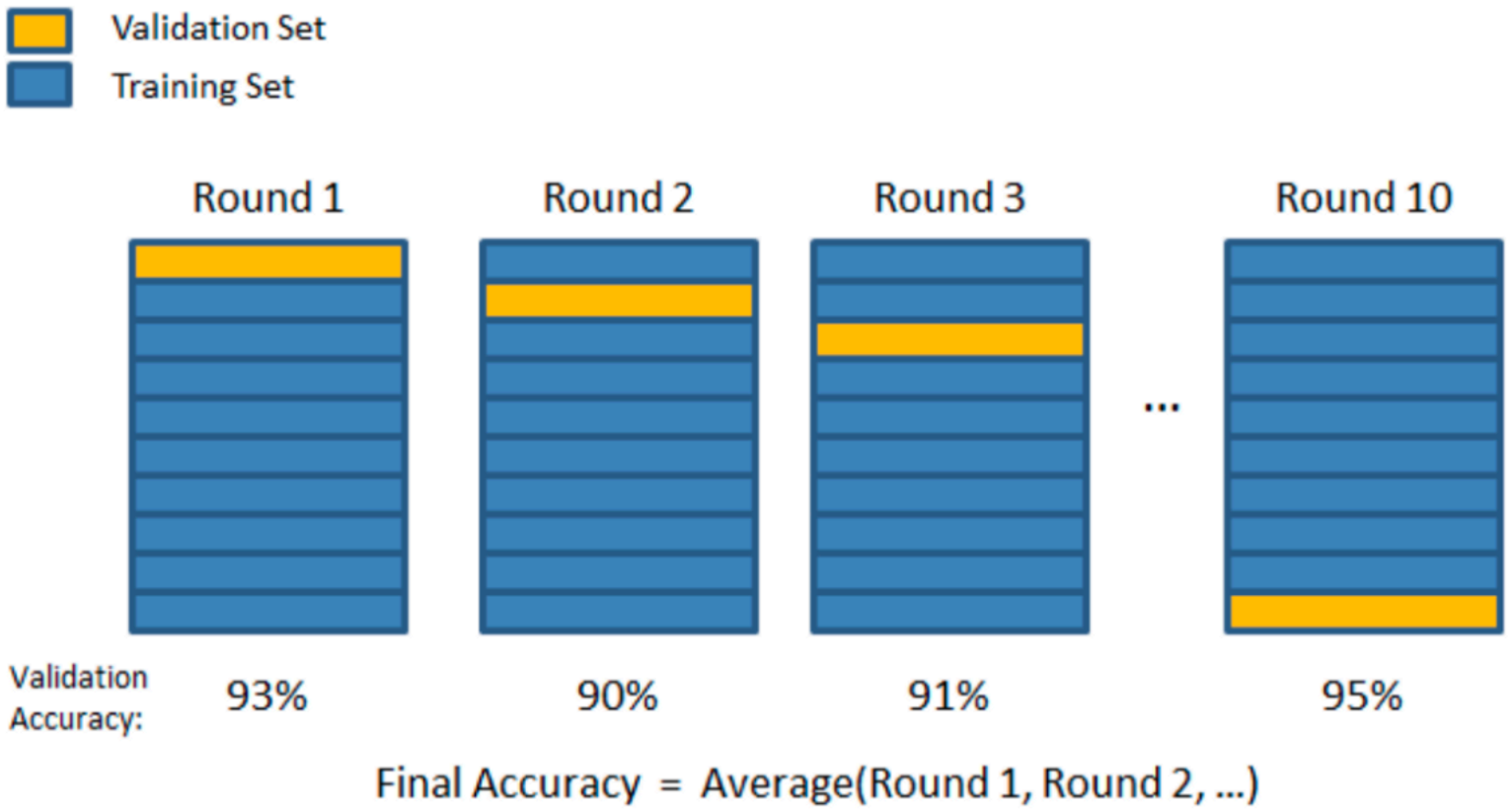


Cross Validation



Joseph Nelson @josephofiowa

K-Folds Cross Validation

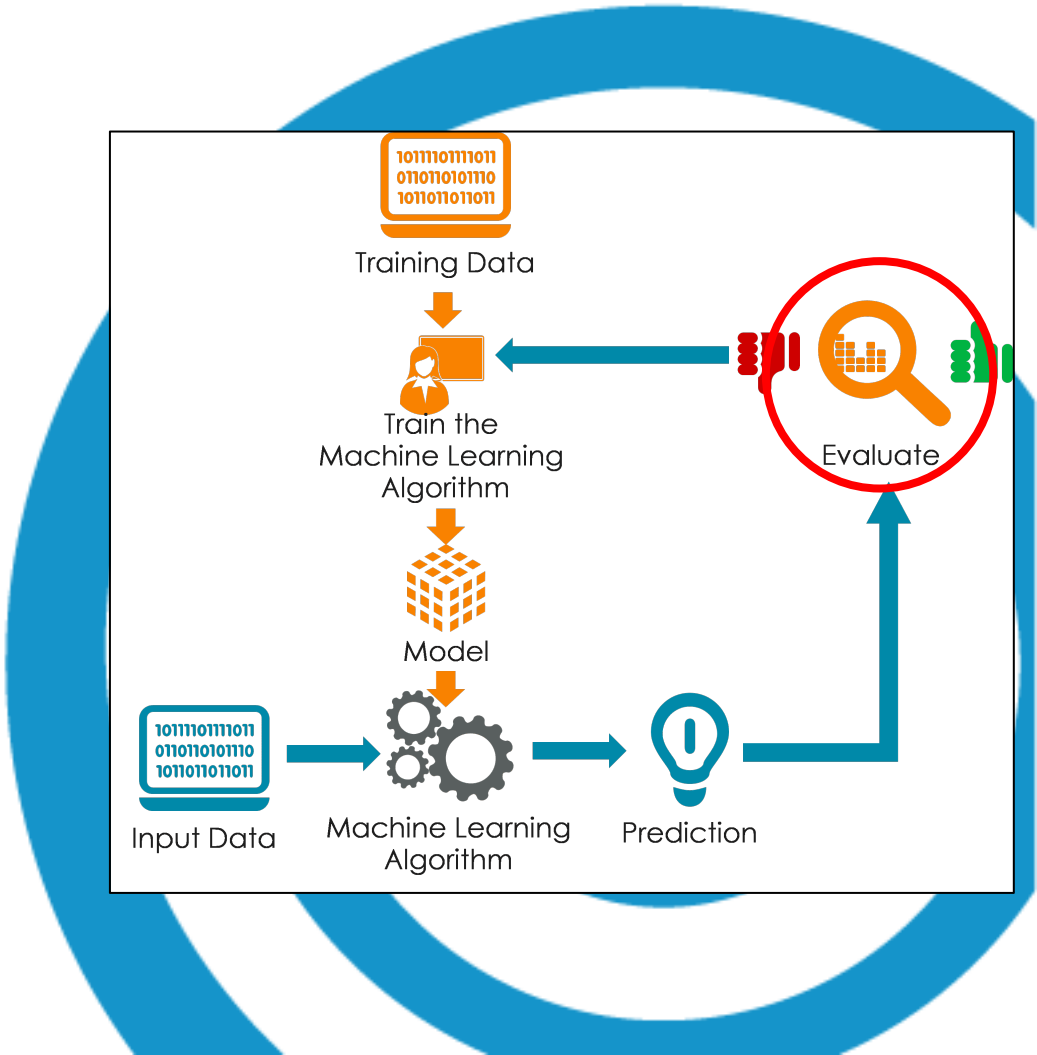


Step 5 Evaluation

Accuracy

ตัวละครในมังกรหยก	ตอบ	จริง	ถูก
เอี้ยก้วย	Yes	Yes	True
อึ้งย้ง	Yes	Yes	True
เซียวเล่งนึ่ง	Yes	Yes	True
จิวแป๊ะทง	Yes	Yes	True
ก๊วยเจ๋ง	Yes	Yes	True
แฮรี่ พอตเตอร์	Yes	No	False

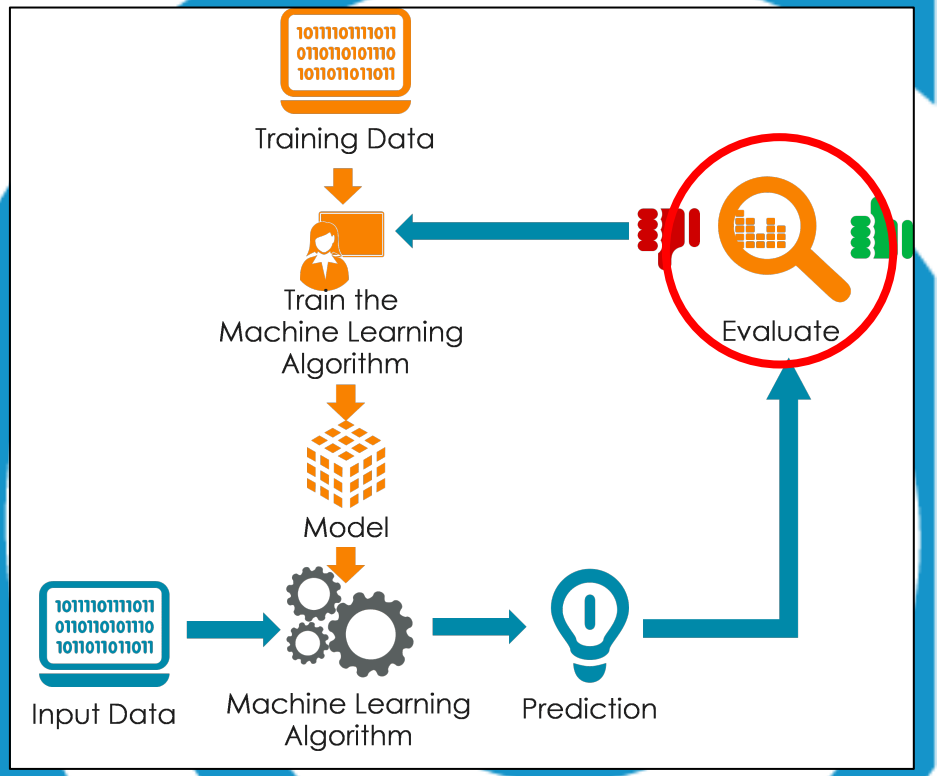
Accuracy = 5/6 = 0.833



Accuracy

ใครติดโควิด-19	ตอบ	จริง	ถูก
เอี้ยก้วย	ไม่ติด	ไม่ติด	True
อึ้งย้ง	ไม่ติด	ไม่ติด	True
เซียวเล่งนึ่ง	ไม่ติด	ไม่ติด	True
จิวแป๊ะทง	ไม่ติด	ไม่ติด	True
ก๊วยเจ๋ง	ไม่ติด	ไม่ติด	True
แฮรี่ พอตเตอร์	ไม่ติด	ติด	False

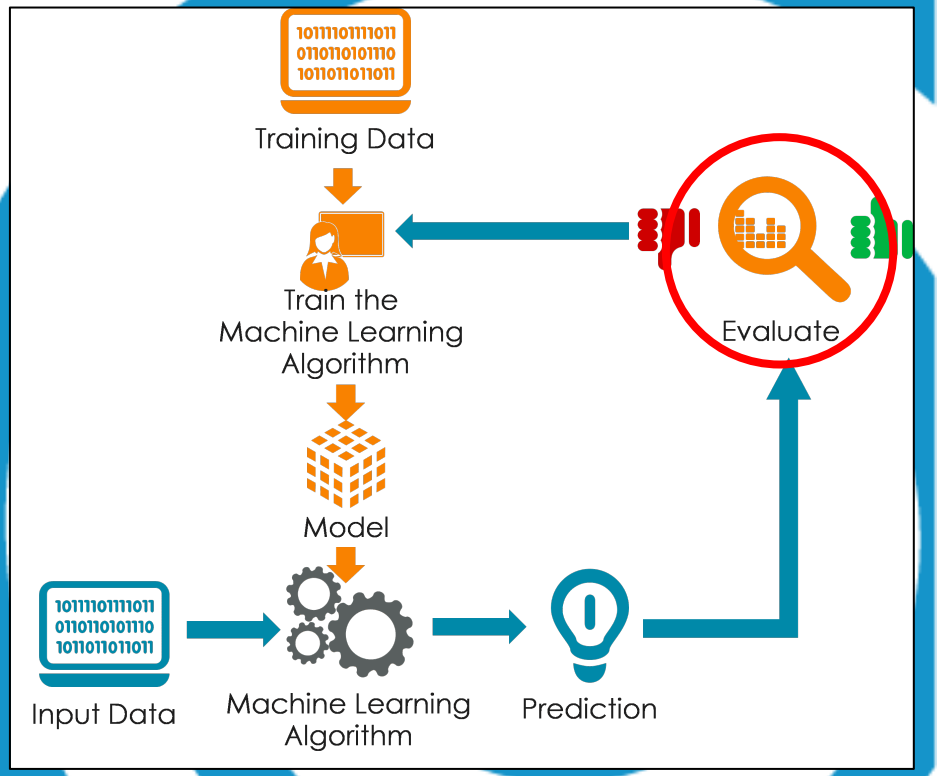
Accuracy = 5/6 = 0.833



Confusion Matrix

Predicted Class

Actual Class	**Positive**	**Negative**	
Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
	Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$



Accuracy

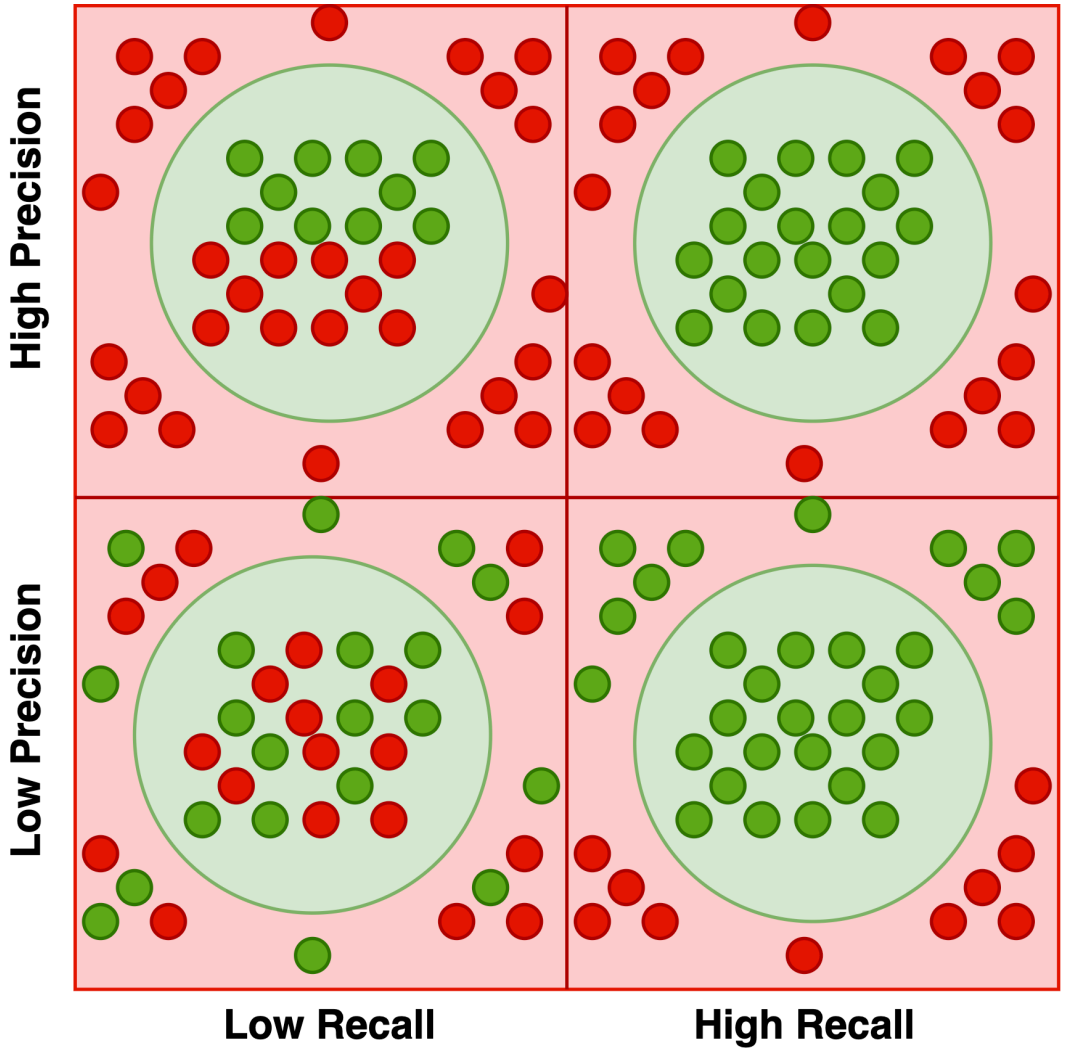
ใครติดโควิด-19	ตอบ	จริง	ถูก
เอี้ยก้วย	ไม่ติด	ไม่ติด	True
อึ้งย้ง	ไม่ติด	ไม่ติด	True
เซียวเล่งนึ่ง	ไม่ติด	ไม่ติด	True
จิวแป๊ะทง	ไม่ติด	ไม่ติด	True
ก๊วยเจ๋ง	ไม่ติด	ไม่ติด	True
แฮรี่ พอตเตอร์	ไม่ติด	ติด	False

Accuracy = 5/6 = 0.833

		Predicted Class: Positive	Predicted Class: Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

5	0
1	0

Precision & Recall



- True Positive
- True Negative
- False Positive
- False Negative

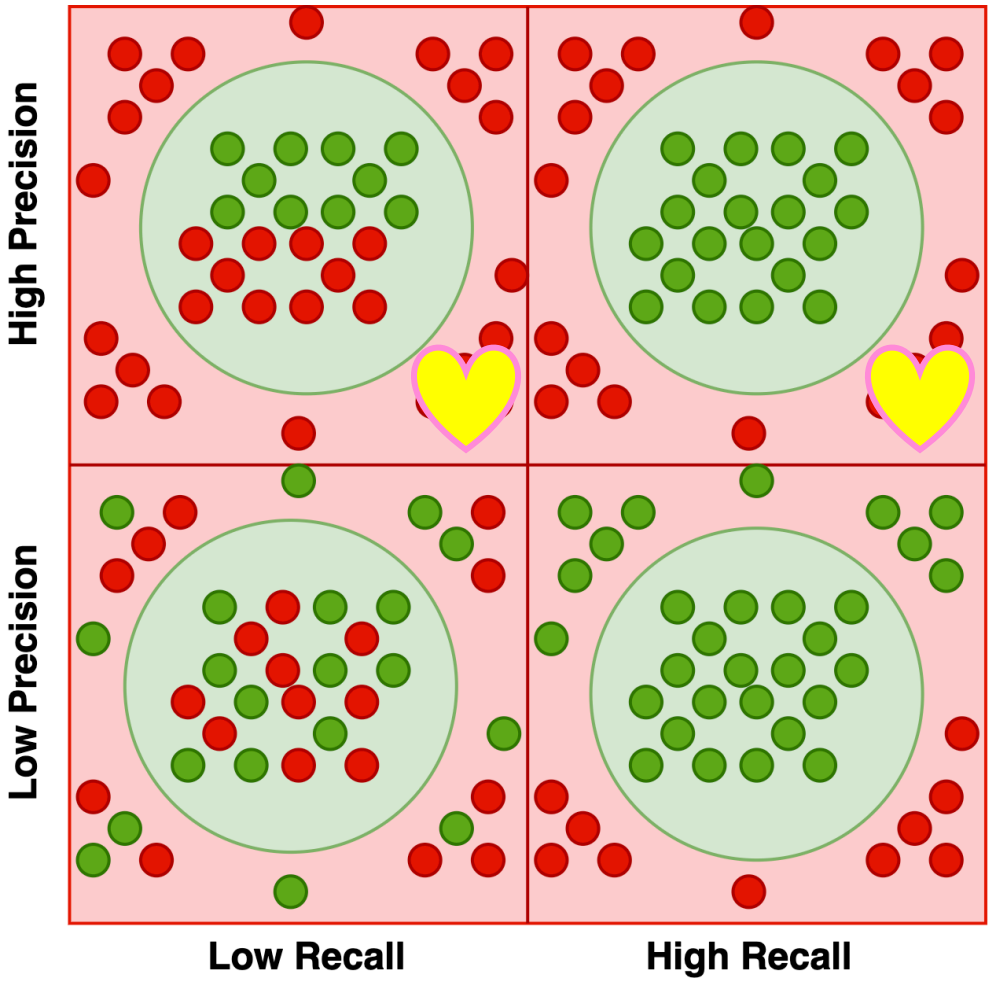
$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision & Recall

Which email is a spam

- Spam
- Not Spam



- True Positive
- True Negative
- False Positive
- False Negative

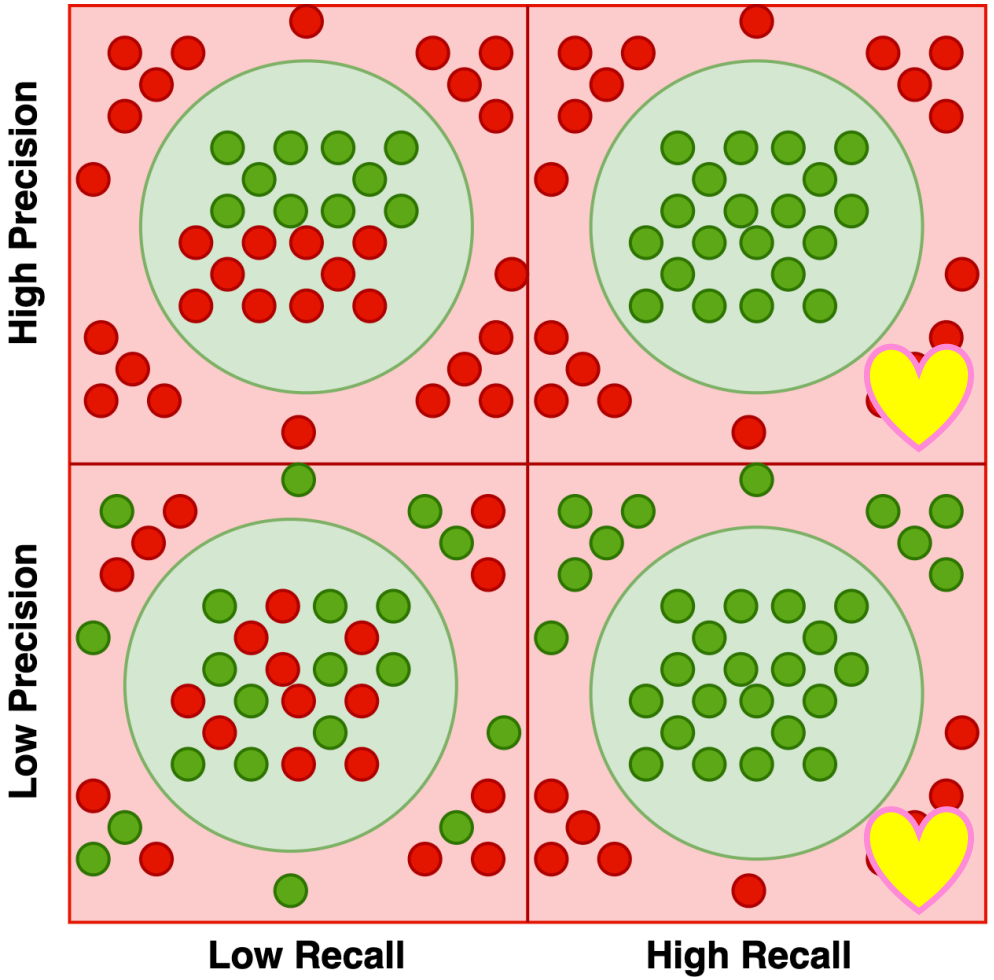
$$\text{Precision} = \frac{TP}{TP + FP}$$

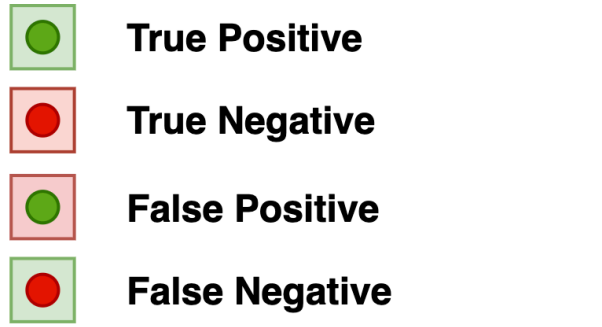
$$\text{Recall} = \frac{TP}{TP + FN}$$

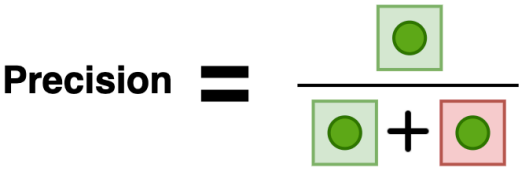
Precision & Recall

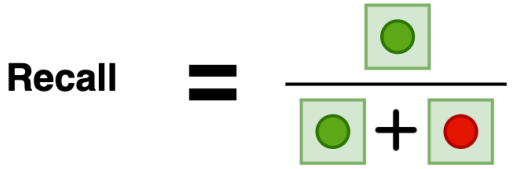
Who is getting infected covid-19

- Infect
- Not Infect



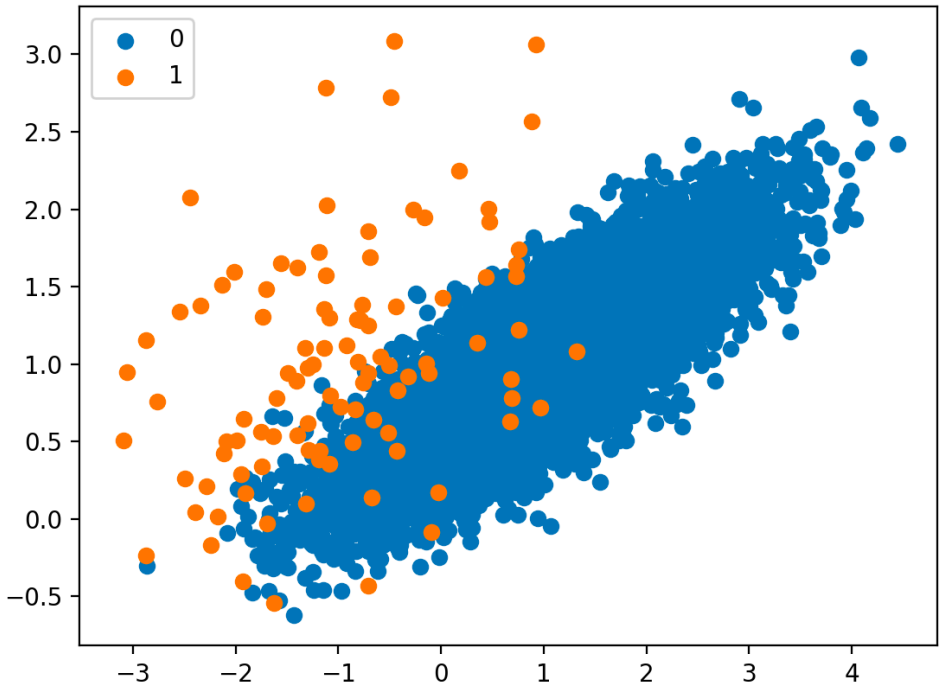




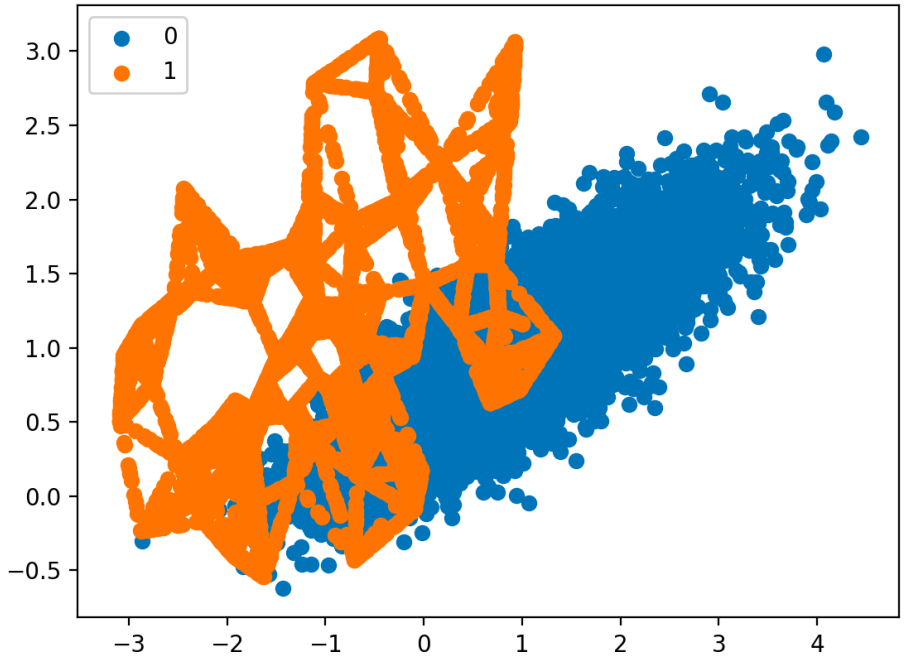


Supervised Learning and Imbalance data

Oversampling



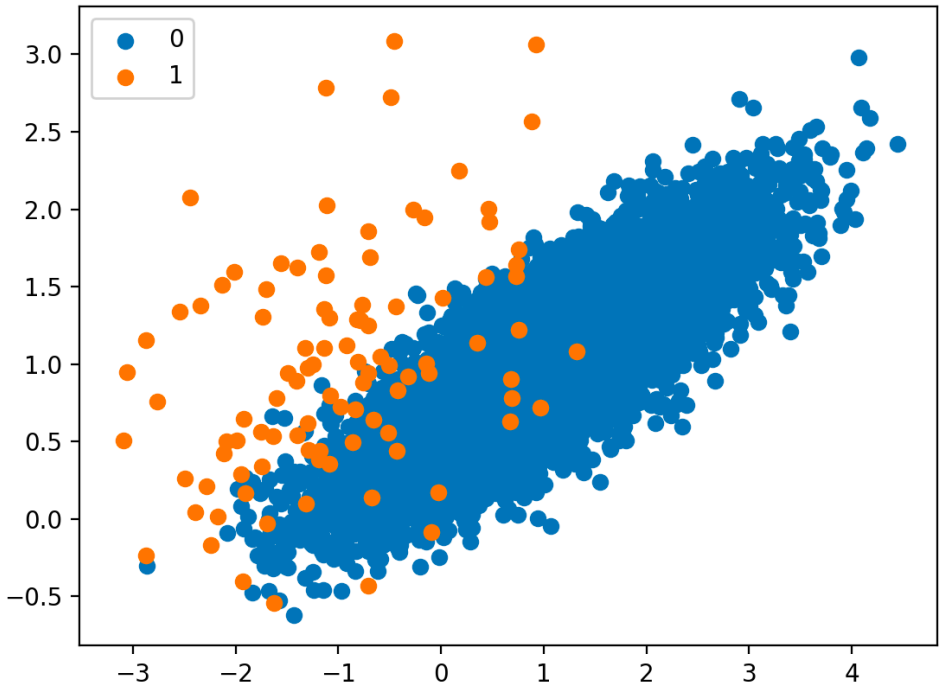
Original dataset



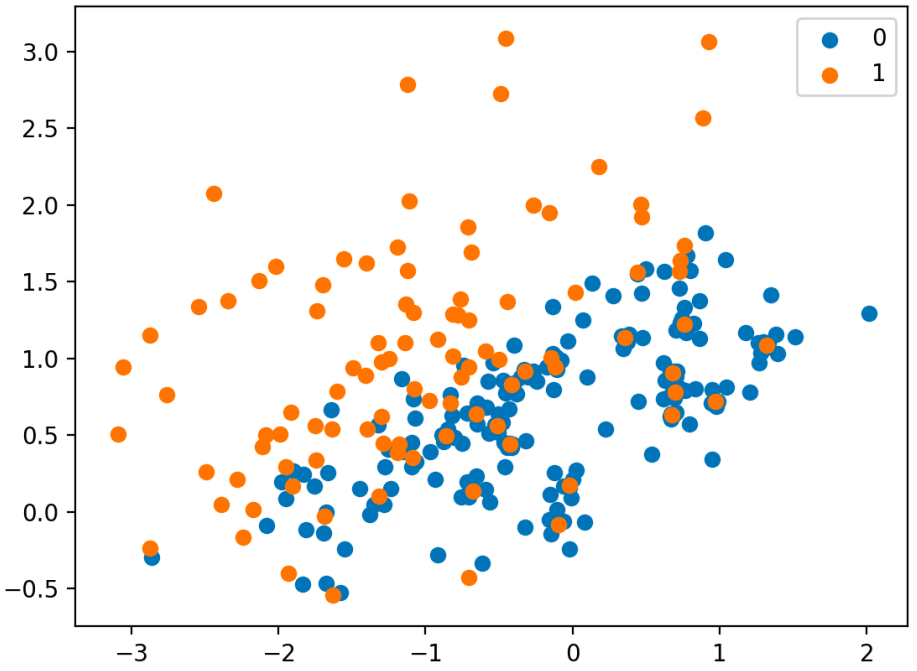
Over sampling with smote

Supervised Learning and Imbalance data

Undersampling



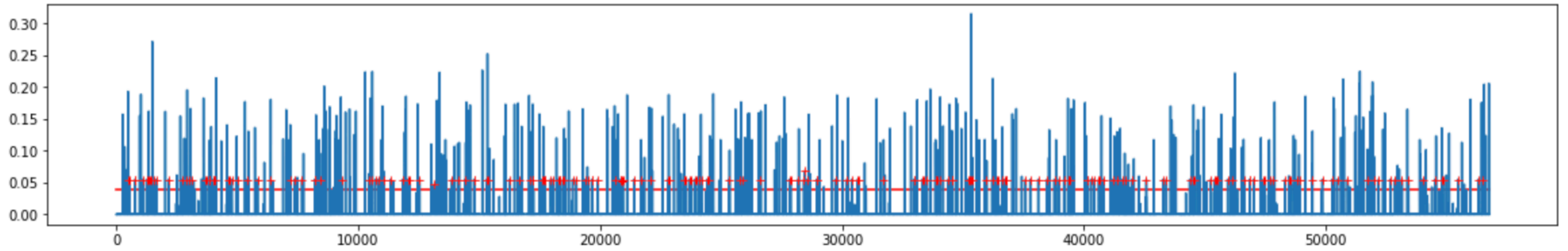
Original dataset



Undersampling

Precision & Recall in Real World

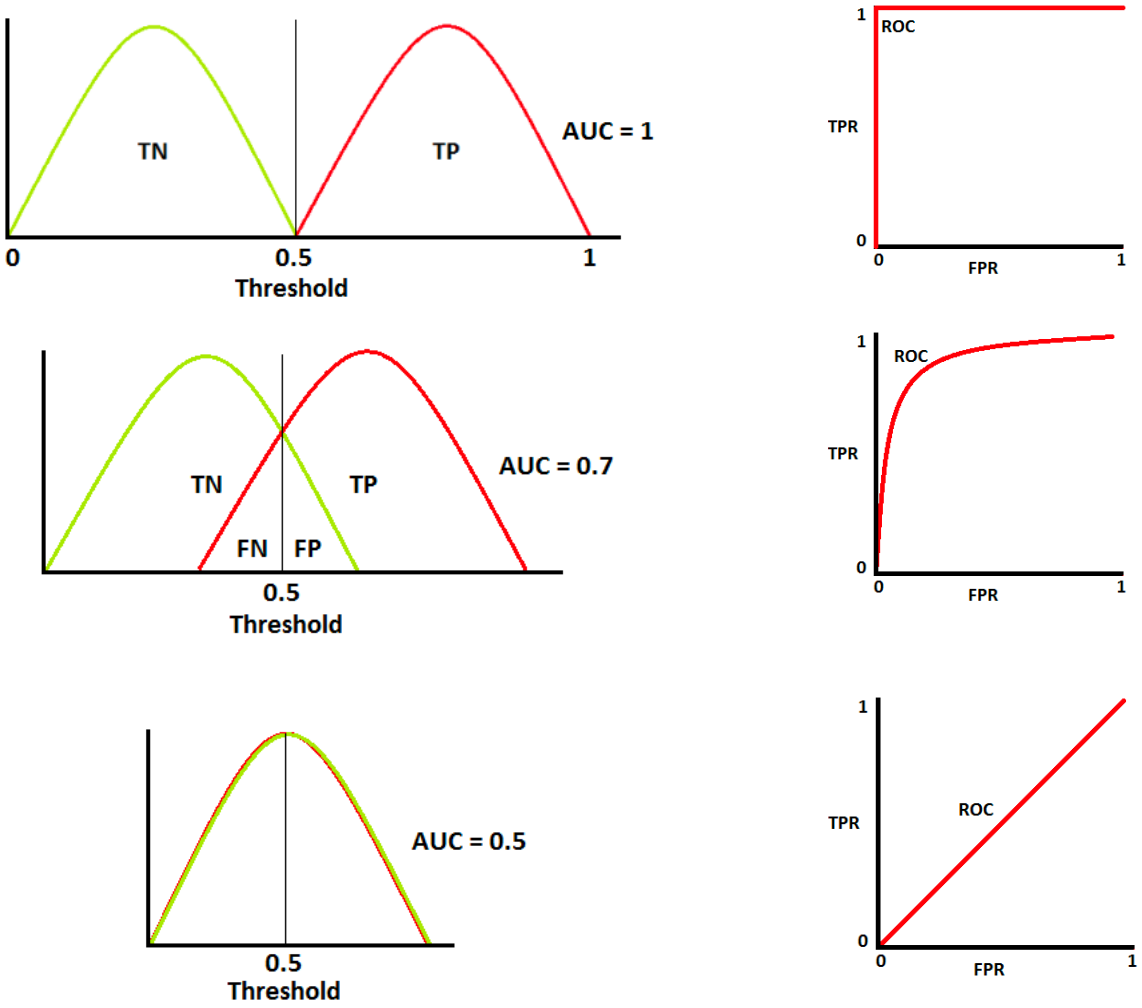
	precision	recall	f1-score	support
0	0.99	1.00	1.00	56225
1	1.00	0.40	0.57	525
accuracy			0.99	56750
macro avg	1.00	0.70	0.78	56750
weighted avg	0.99	0.99	0.99	56750



เมื่อกำหนด threshold ที่ 0.04 พบว่า

- recall ของ anomaly เท่ากับ 0.4 หมายความว่าเรา Pred เป็น Anomaly ใน 100% เรา Pred ถูก 40%
- precision ของ anomaly เท่ากับ 1 หมายความว่า ใน 40% จาก recall เรา Pred ถูกหมด 100%

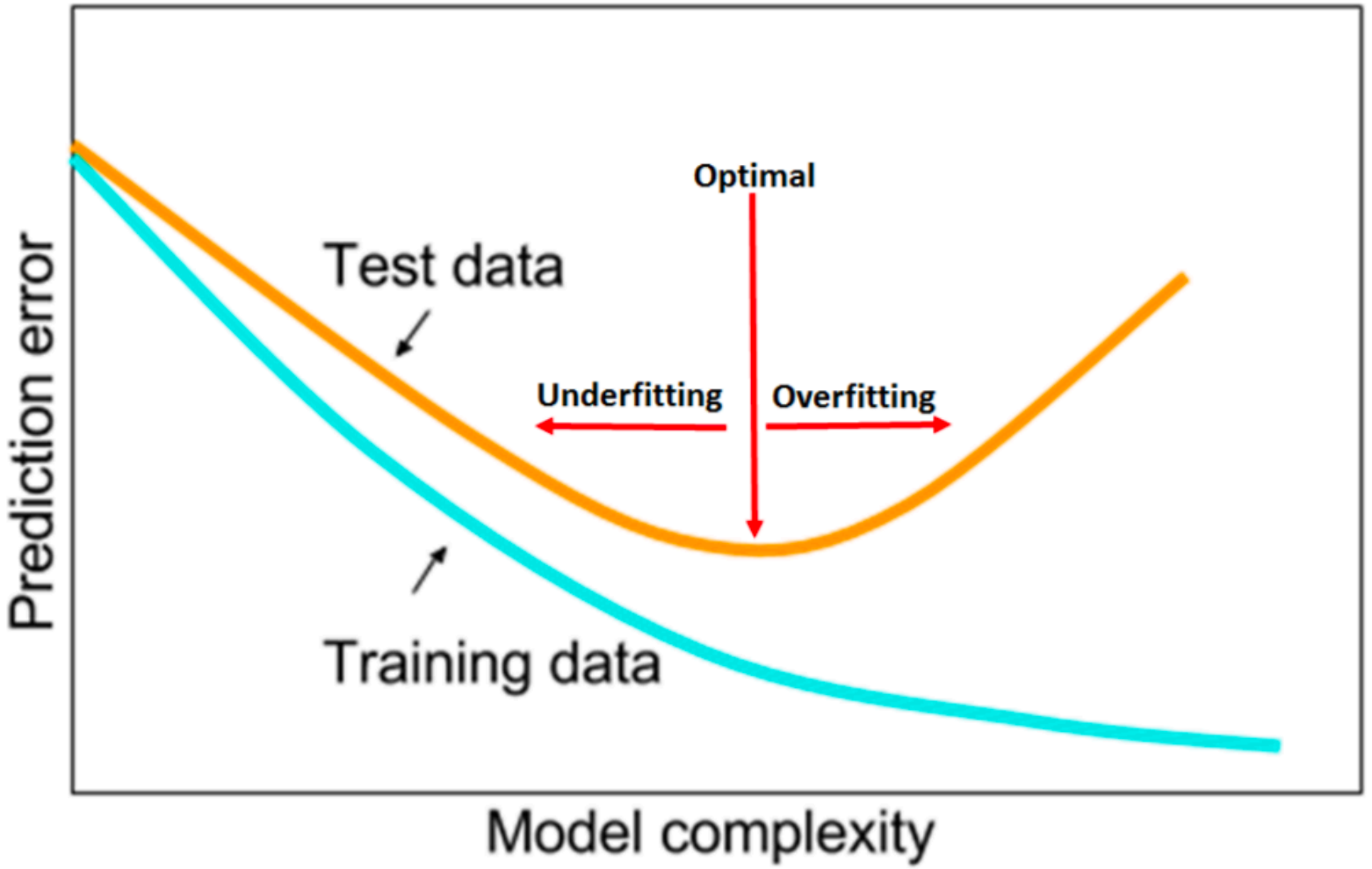
ROC and AUC



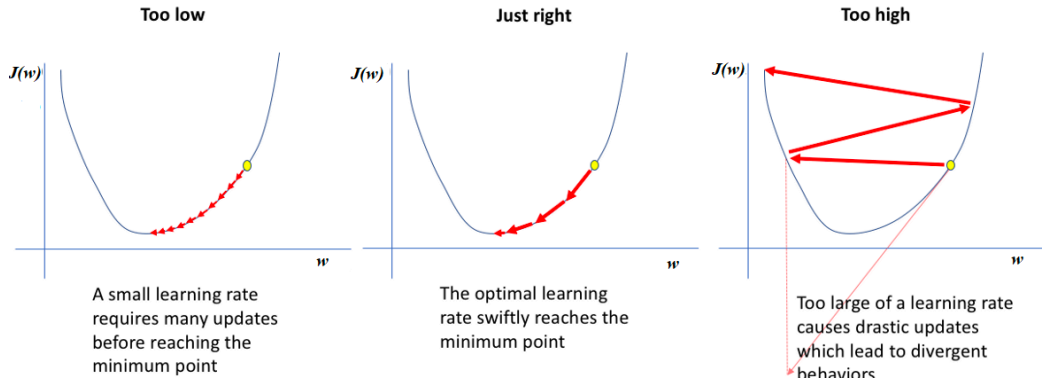
https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

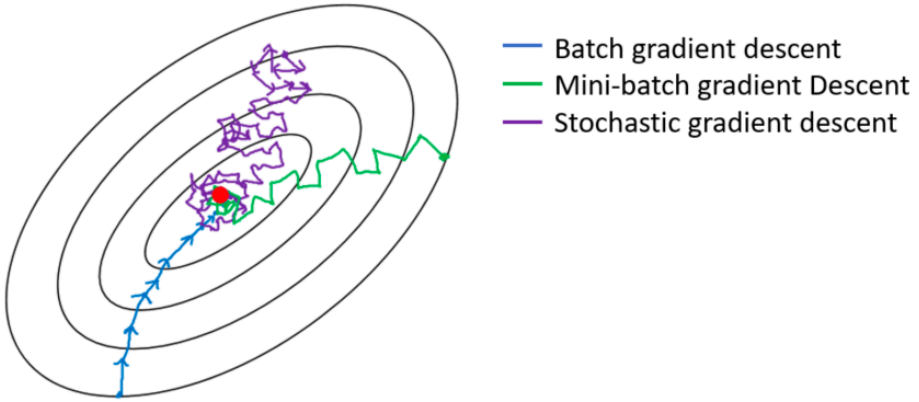
Step 6-7 Parameter tuning & Prediction

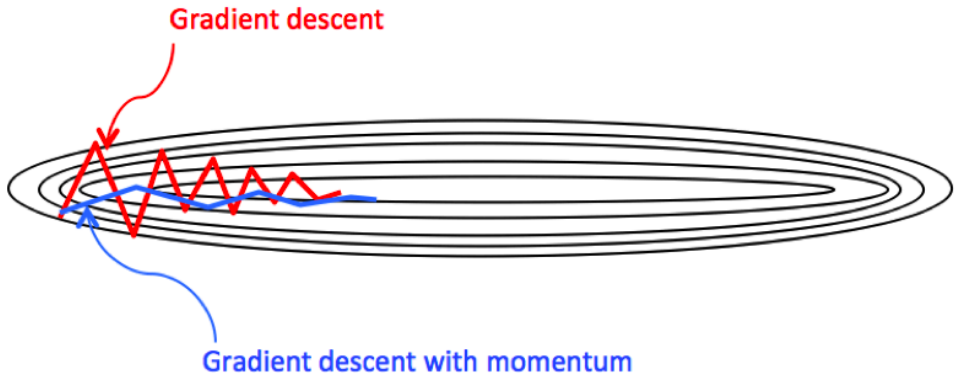
Overfit


Optimal
Test data
Underfitting
Overfitting
Training data
Prediction error
Model complexity

Hyper–parameter Tuning

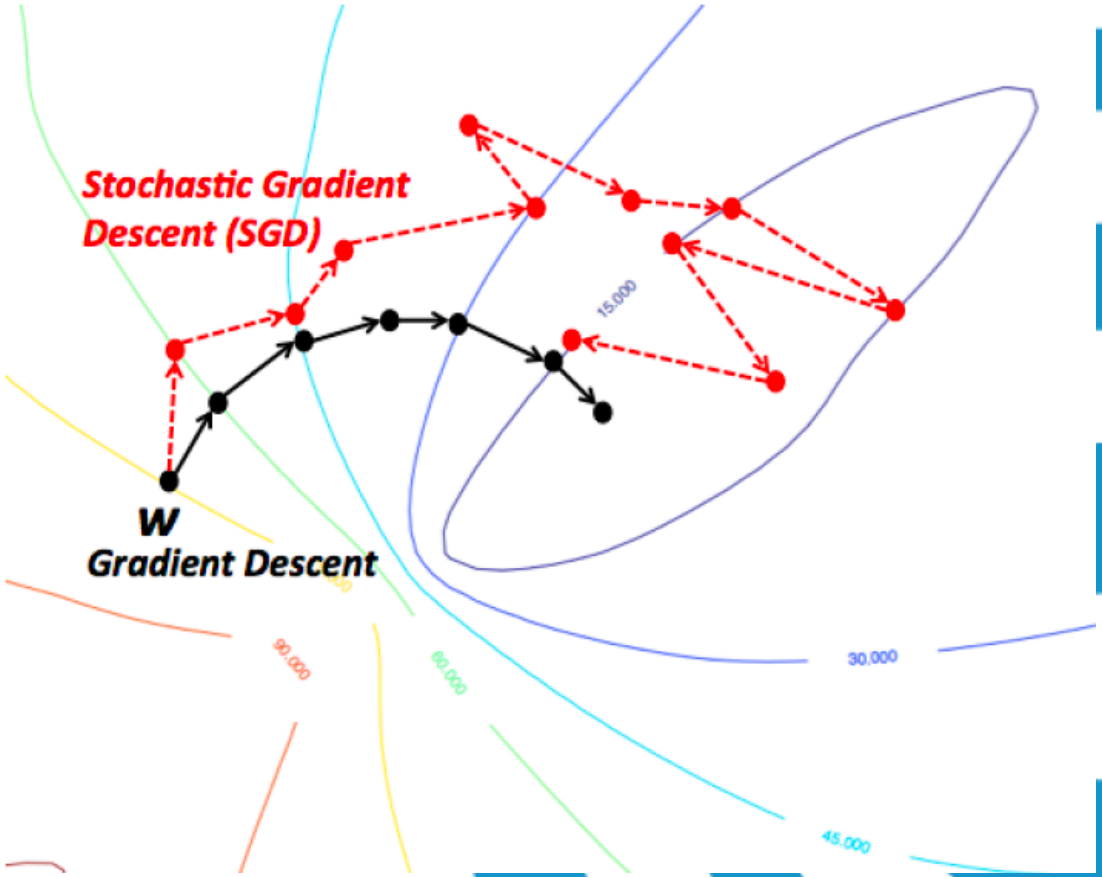






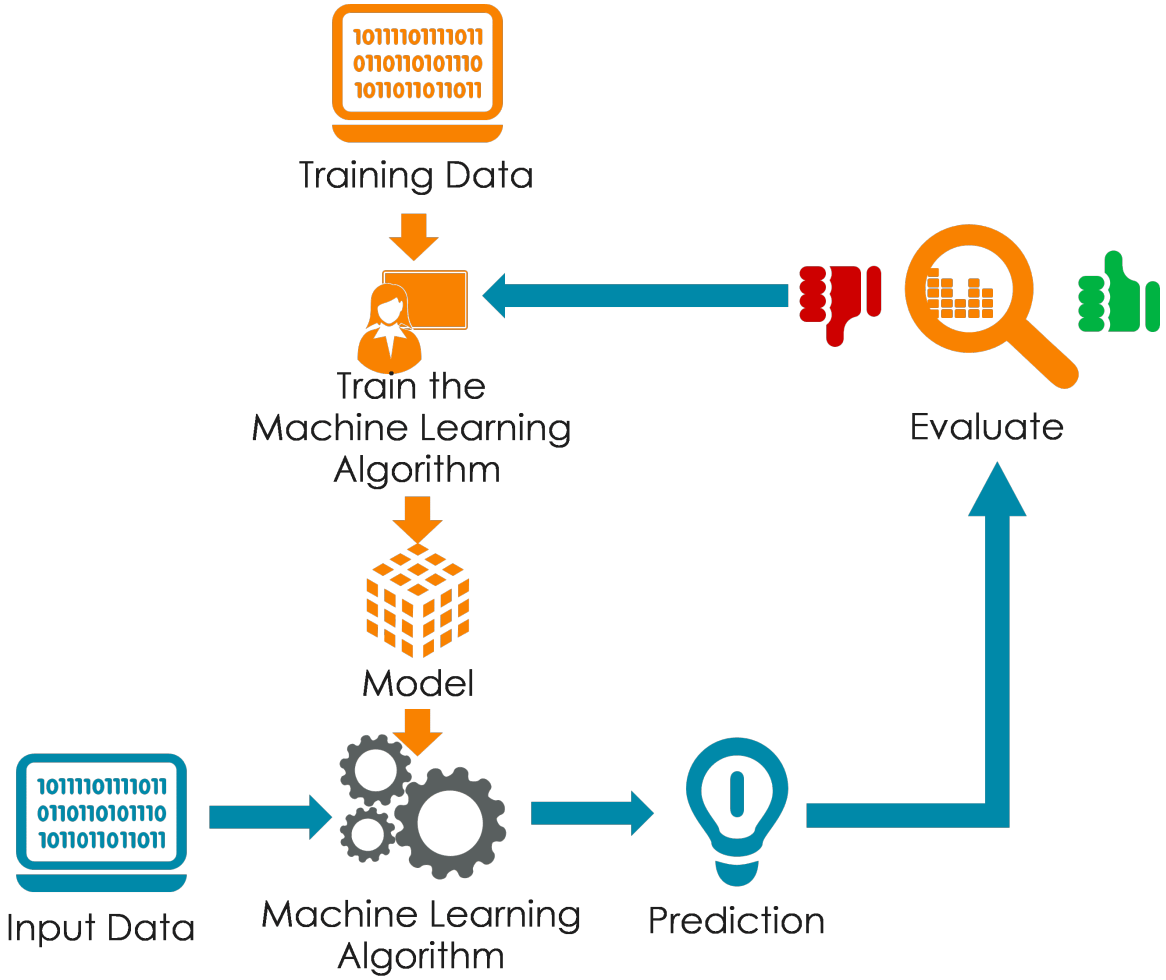
Optimizers

- SGD
- RMSprop
- Adam
- Adadelta
- Adagrad
- Adamax
- Nadam
- Ftrl



https://keras.io/api/optimizers/

Prediction


Training Data
Train the Machine Learning Algorithm
Evaluate
Model
Input Data
Machine Learning Algorithm
Prediction

The Machine Learning Life Cycle

1. **Define Project Objectives**
   - ☐ Specify business problem
   - ☐ Acquire subject matter expertise
   - ☐ Define unit of analysis and prediction target
   - ☐ Prioritize modeling criteria
   - ☐ Consider risks and success criteria
   - ☐ Decide whether to continue
2. **Acquire & Explore Data**
   - ☐ Find appropriate data
   - ☐ Merge data into single table
   - ☐ Conduct exploratory data analysis
   - ☐ Find and remove any target leakage
   - ☐ Feature engineering
3. **Model Data**
   - ☐ Variable selection
   - ☐ Build candidate models
   - ☐ Model validation and selection
4. **Interpret & Communicate**
   - ☐ Interpret model
   - ☐ Communicate model insights
5. **Implement, Document & Maintain**
   - ☐ Set up batch or API prediction system
   - ☐ Document modeling process for reproducibility
   - ☐ Create model monitoring and maintenance plan

https://www.datarobot.com/wiki/machine-learning-life-cycle/

Softnix

Thankyou