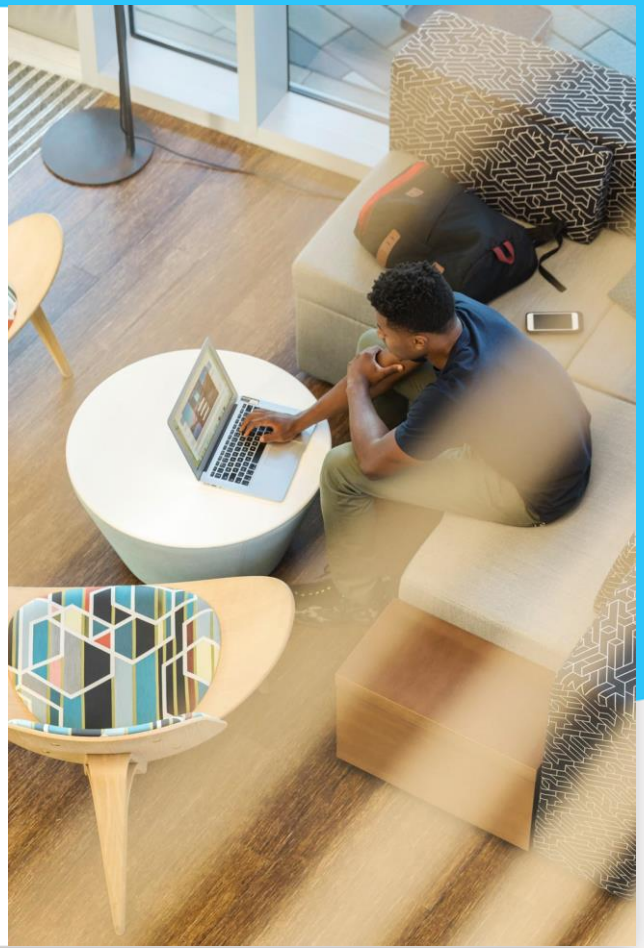
Introduction to Machine Learning

MACHINE LEARNING

Data

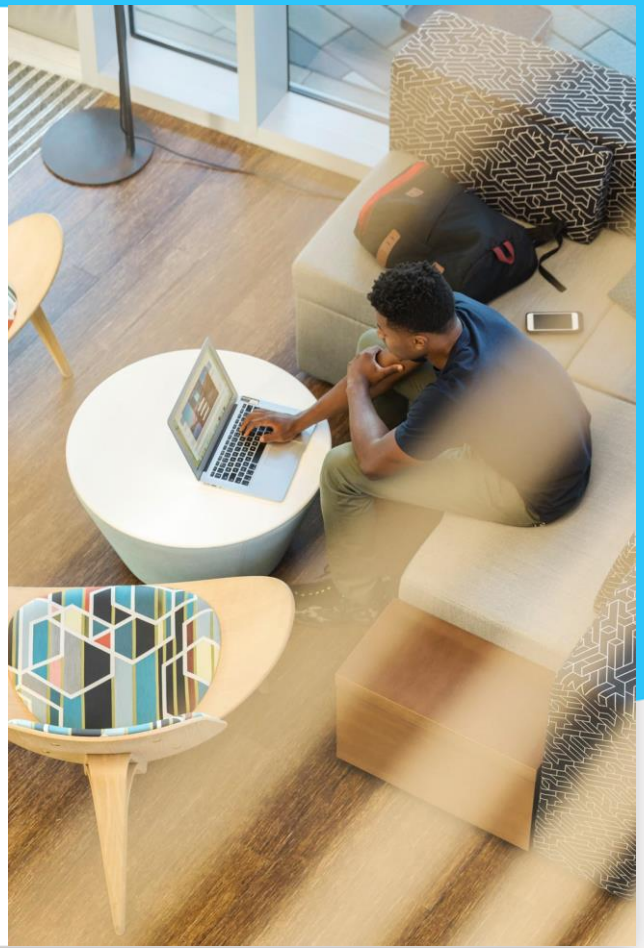
- Continuous process
- Measures
- Targets
- Monitoring



MACHINE LEARNING

Features

- Signal
- Relationships
- Patterns
- Hidden information in our dataset
- Valuable resources to predict our target



MACHINE LEARNING

Learning

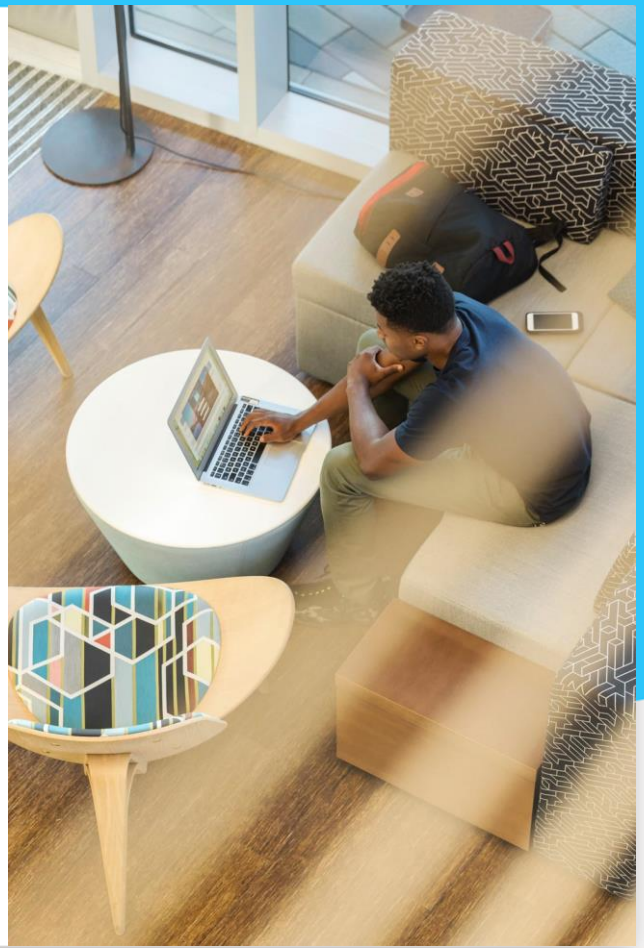
Set of Rules – Human vs Machine

Learn from Data

*.Learn a pattern so that when it sees **similar** data, it will be able to understand it.*

.Take data as input, and output a prediction

Data → Machine Learning → **Output**



MACHINE LEARNING

Techniques

Supervised Learning

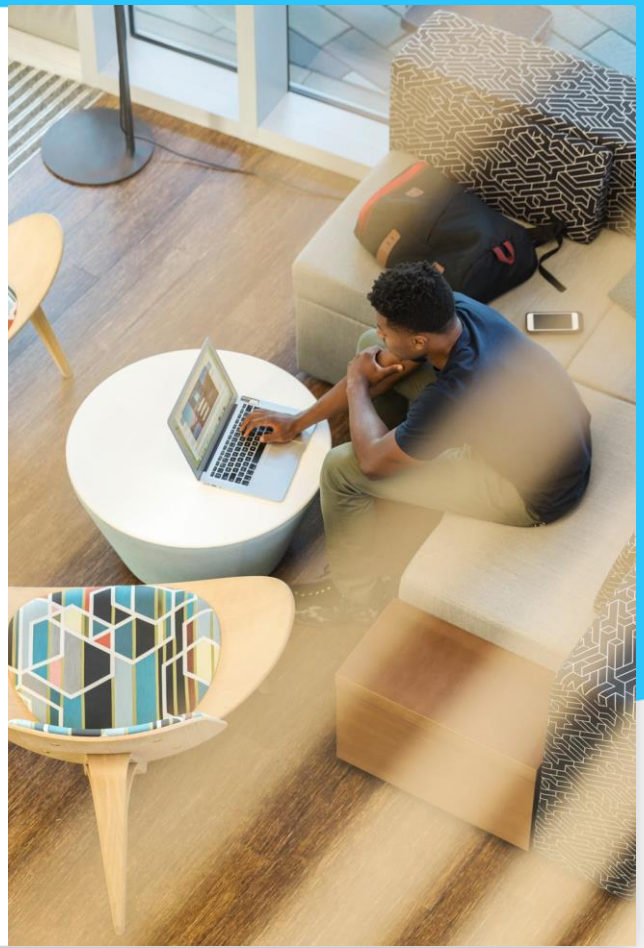
. *Regression*

. *Classification*

Unsupervised Learning

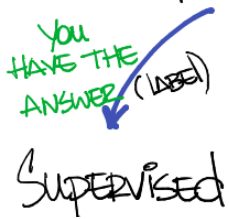
. *Dimensionality Reduction*

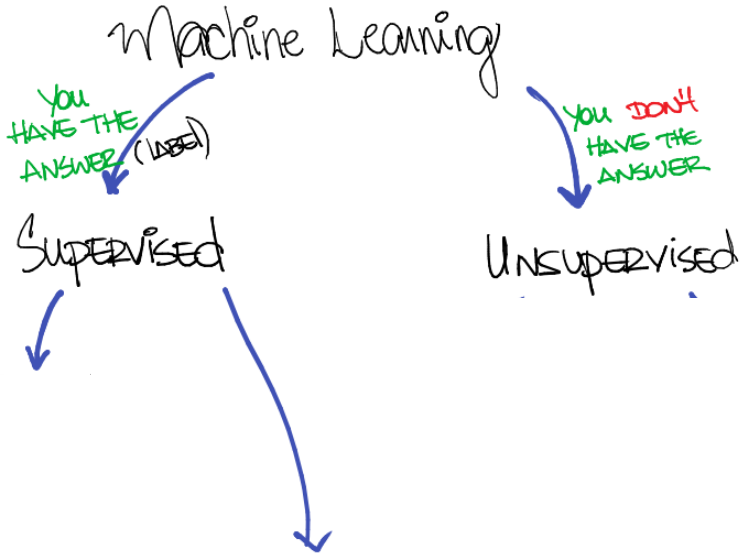
. *Clustering*



Machine Learning

Machine Learning




Machine Learning
You HAVE THE ANSWER (LABEL)
You DON'T HAVE THE ANSWER
Supervised
Unsupervised

Machine Learning

You have the answer (label)

You don't have the answer

Supervised

Unsupervised

Regression

Predict a real number

Machine Learning

You HAVE THE ANSWER (LABEL)

You DON'T HAVE THE ANSWER

Supervised

Regression

Predict a real number

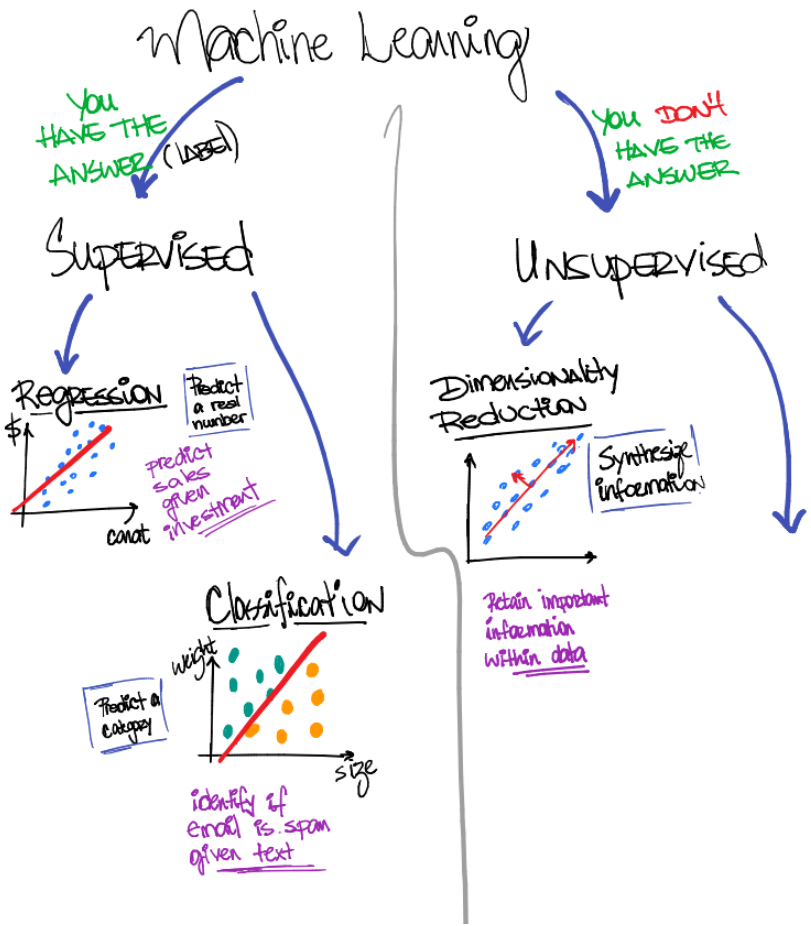
predict sales given investment

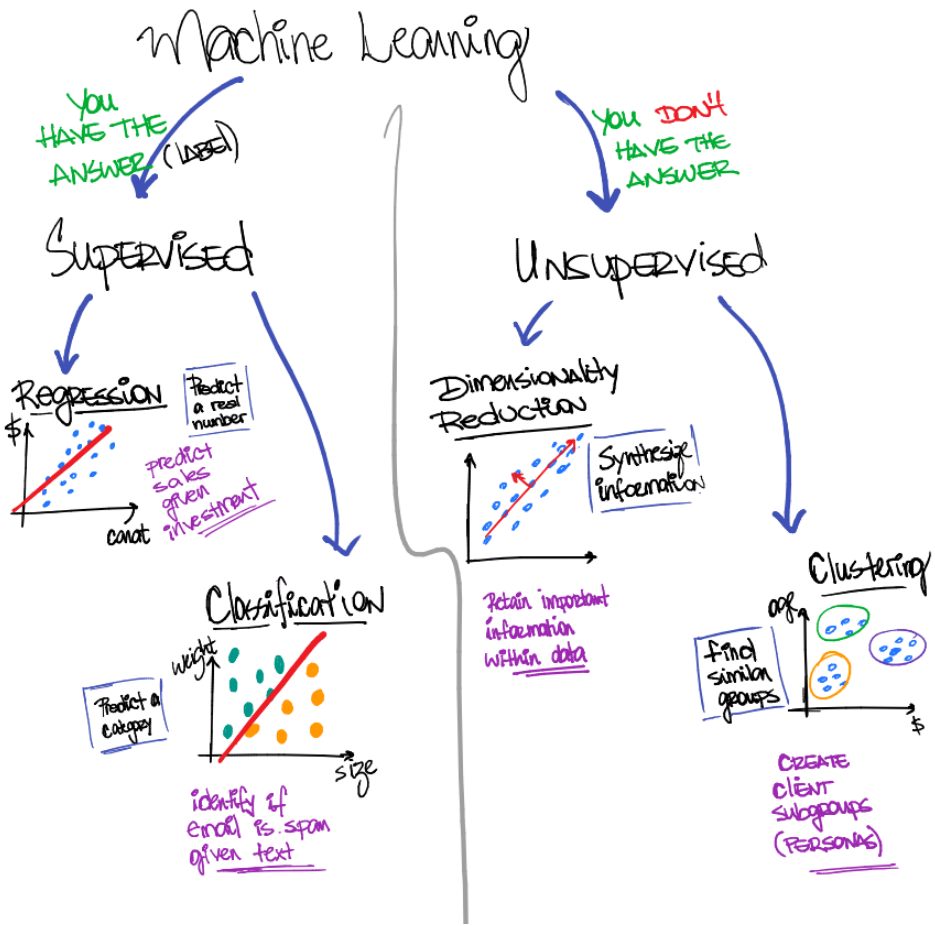
Classification

Predict a category

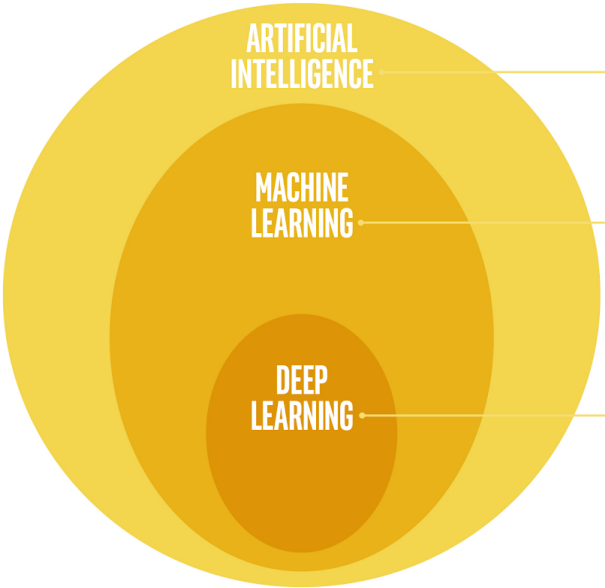
identify if email is spam given text

Unsupervised

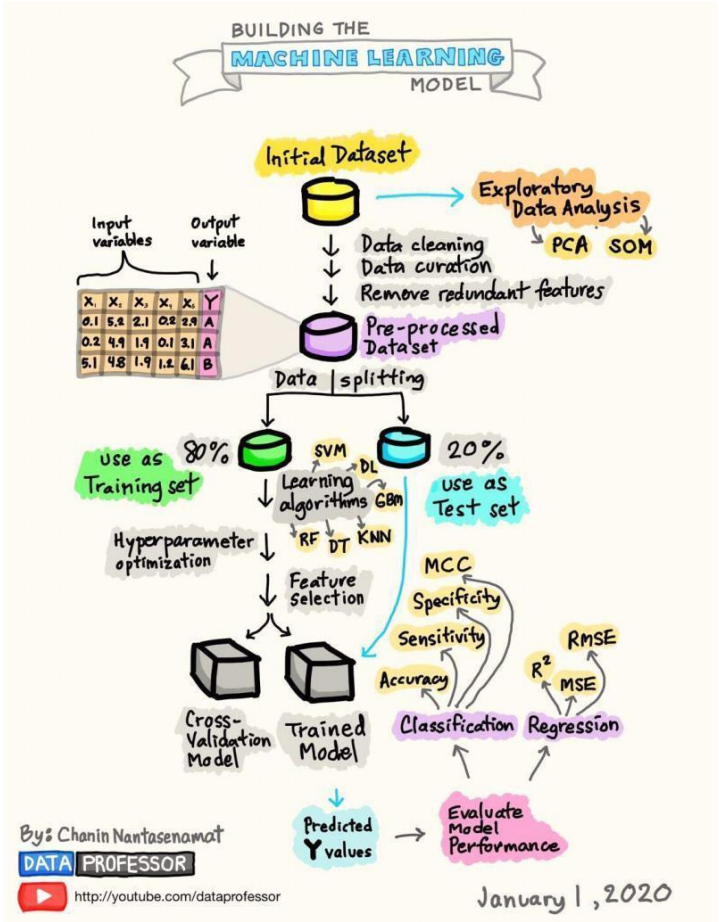

Machine Learning
You HAVE THE ANSWER (LABEL)
You DON'T HAVE THE ANSWER
Supervised
Unsupervised
Regression
Predict a real number
$
cost
predict sales given investment
Classification
Predict a category
weight
size
identify if email is spam given text
Dimensionality Reduction
Synthesize information
Retain important information within data


Machine Learning
You HAVE THE ANSWER (LABEL)
You DON'T HAVE THE ANSWER
Supervised
Unsupervised
Regression
Predict a real number
$
cost
predict sales given investment
Classification
weight
size
Predict a category
identify if email is spam given text
Dimensionality Reduction
Synthesize information
Retain important information within data
Clustering
age
$
find similar groups
CREATE CLIENT SUBGROUPS (PERSONAS)

ARTIFICIAL INTELLIGENCE TERMS



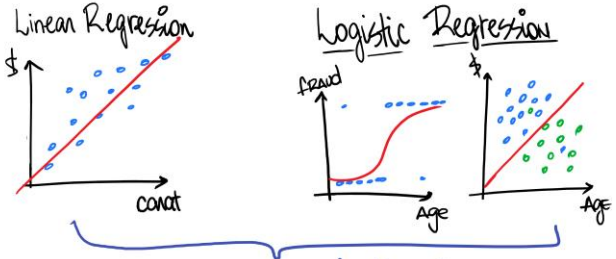
- **AI** is an umbrella term for machines capable of perception, logic, and learning.
- **Machine learning** employs algorithms that learn from data to make predictions or decisions, and whose performance improves when exposed to more data over time.
- **Deep learning** uses many-layered neural networks to build algorithms that find the best way to perform tasks on their own, based on vast sets of data.


BUILDING THE
MACHINE LEARNING
MODEL
Initial Dataset
Exploratory Data Analysis
PCA
SOM
Input variables
Output variable
Data cleaning
Data curation
Remove redundant features
Pre-processed Data set
Data splitting
80%
use as Training set
20%
use as Test set
SVM
DL
Learning algorithms
GBM
RF
DT
KNN
Hyperparameter optimization
Feature selection
MCC
Specificity
Sensitivity
Accuracy
RMSE
R^2
MSE
Classification
Regression
Cross-Validation Model
Trained Model
Predicted Y values
Evaluate Model Performance
By: Chanin Nantasenamat
DATA PROFESSOR
http://youtube.com/dataprofessor
January 1, 2020

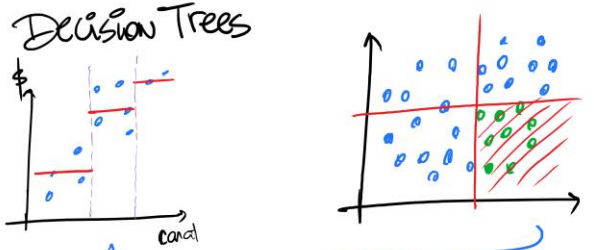
Machine Learning Methods

Regression

Classification

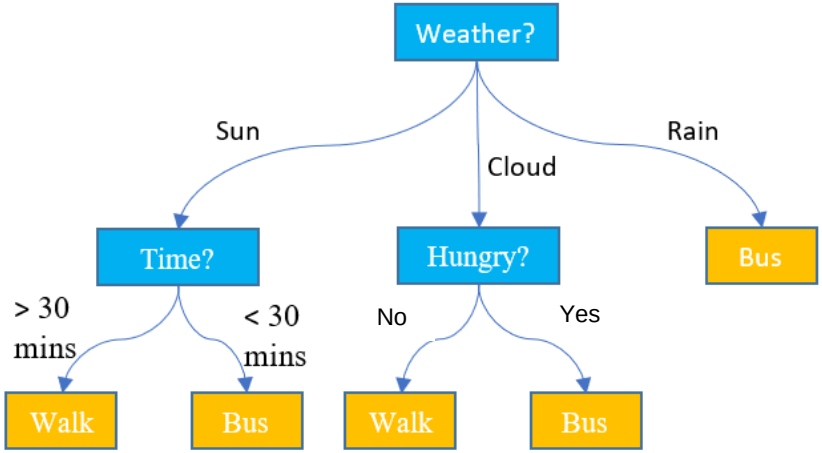


LINEAR METHODS



NON-LINEAR METHODS

Should I go walking or by bus?



Bias and Variance TradeOff

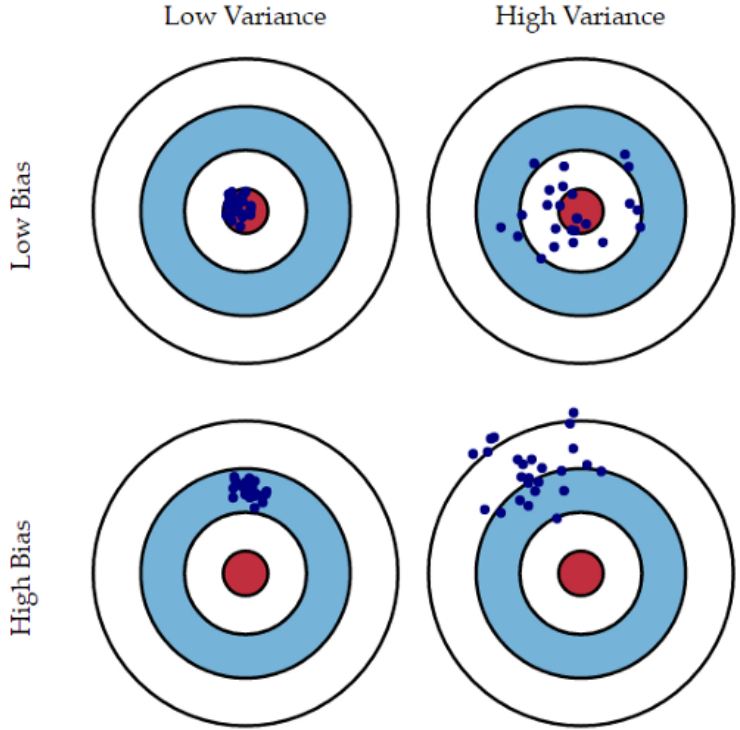
Bias

A highly biased model oversimplifies the information given by your data and tends to have high error rate.

Variance

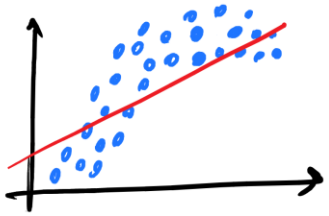
A model with high variance tends to pay so much attention to the data it was given that it fails to generalize for data it hasn’t seen before.

Bias and Variance TradeOff

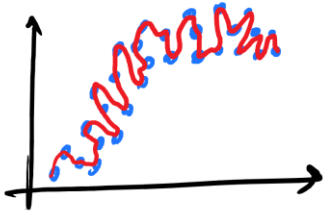


Machine Learning

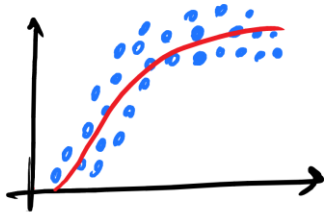
Overfitting : Regression



: under fitting

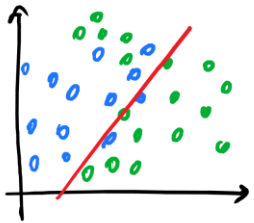


: OVERfitting

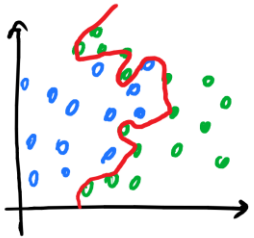


: generalization

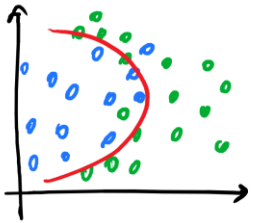
Machine Learning Overfitting : classification



: under fitting

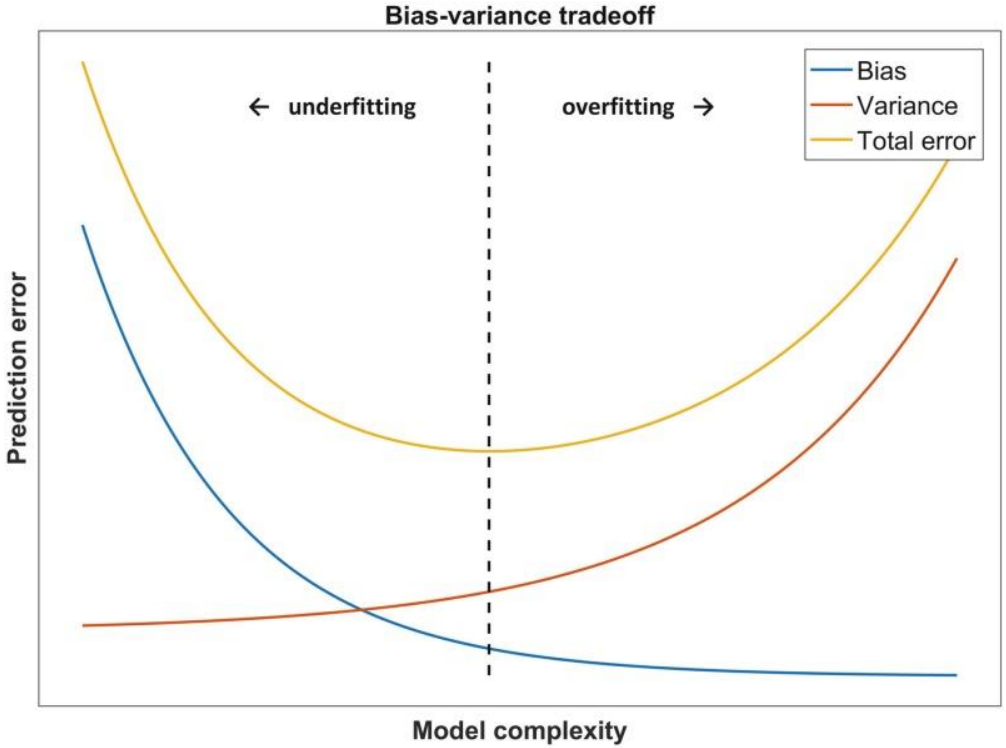


: overfitting

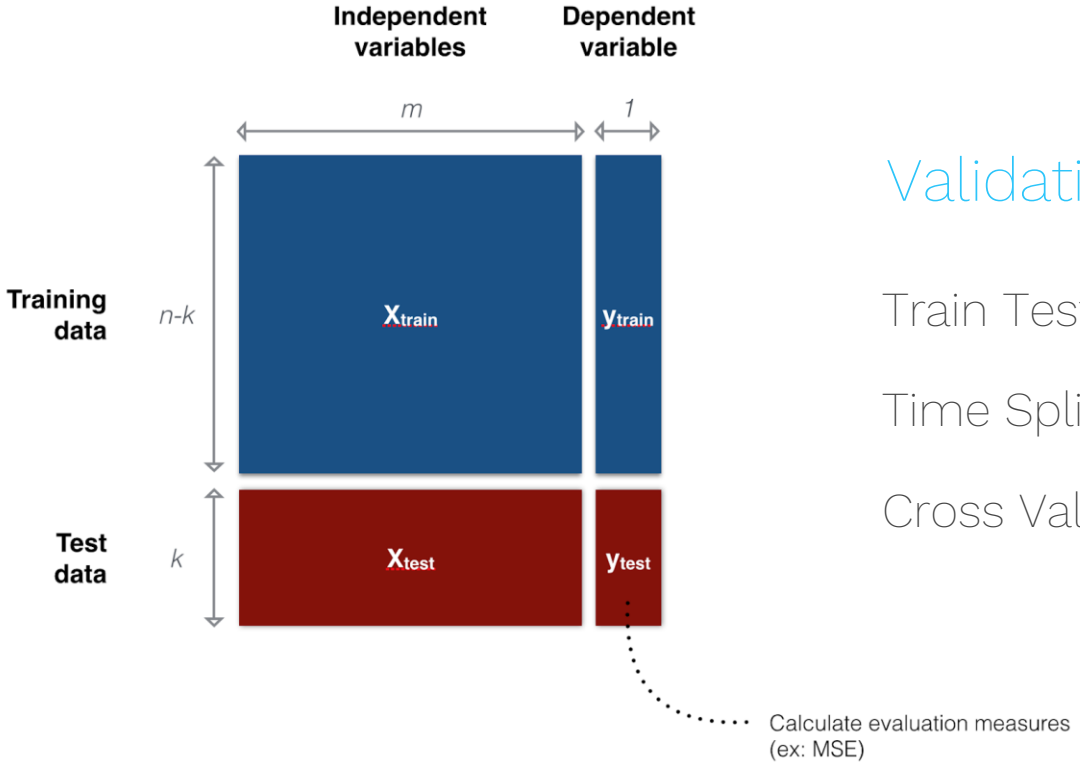


: generalization

Bias and Variance TradeOff



Techniques to prevent overfitting



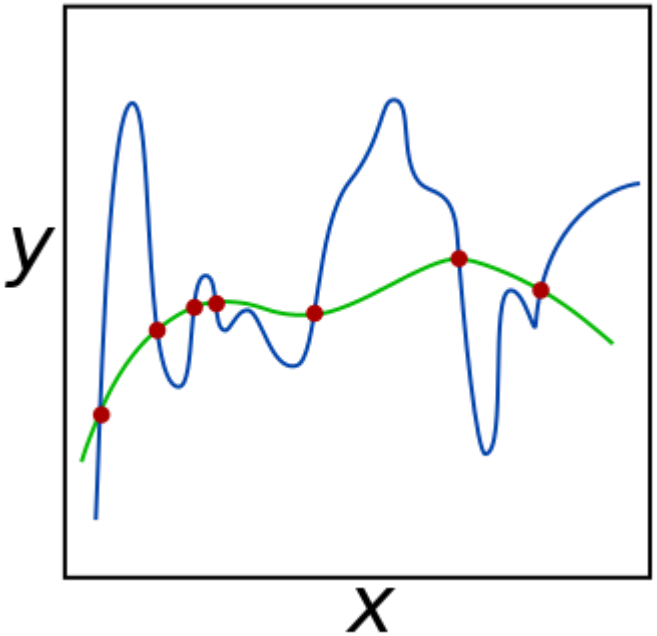
Validation Techniques

Train Test Split

Time Split

Cross Validation

Techniques to prevent overfitting



Reducing Model Complexity

Regularization

Tree Pruning

Ironhack