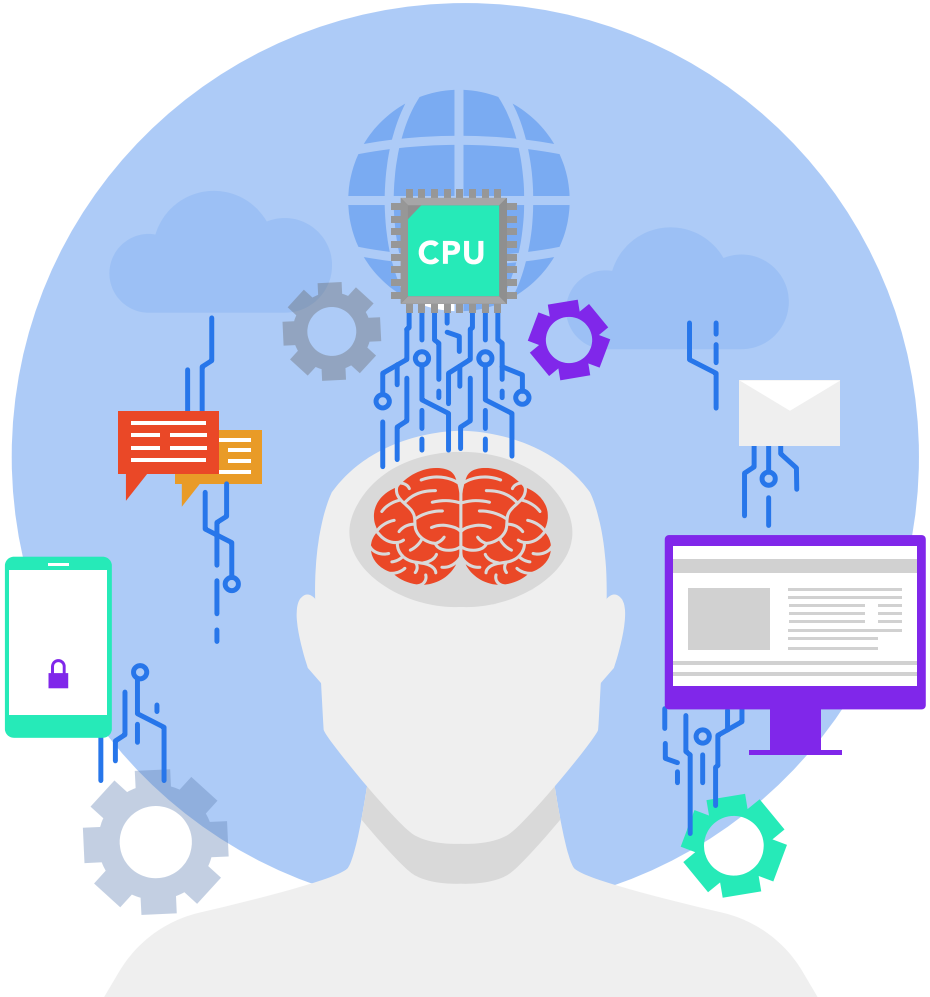


Introduction to Machine Learning

OUTLINE

- Introduction to Machine Learning
- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning
- Reinforcement Learning
- Model Selection and Assessment

Learning Algorithm Definition

Definition

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E” (Mitchell, 1997)

Example

Playing Checkers

E = the experience of playing many games of checkers
T = the task of playing checkers
P = the probability that the program will win the next game

Another Definition

An algorithm that is able to learn from data.

The Task T

- Machine learning tasks are usually described in terms of **how the machine learning system should process an example**.
- An **example** is a collection of **features** that have been **quantitatively measured** from some objects or events that we want the machine learning system to process.
- Some common **tasks**:
  - Classification
  - Regression
  - Machine Translation
  - Anomaly detection

The Task T

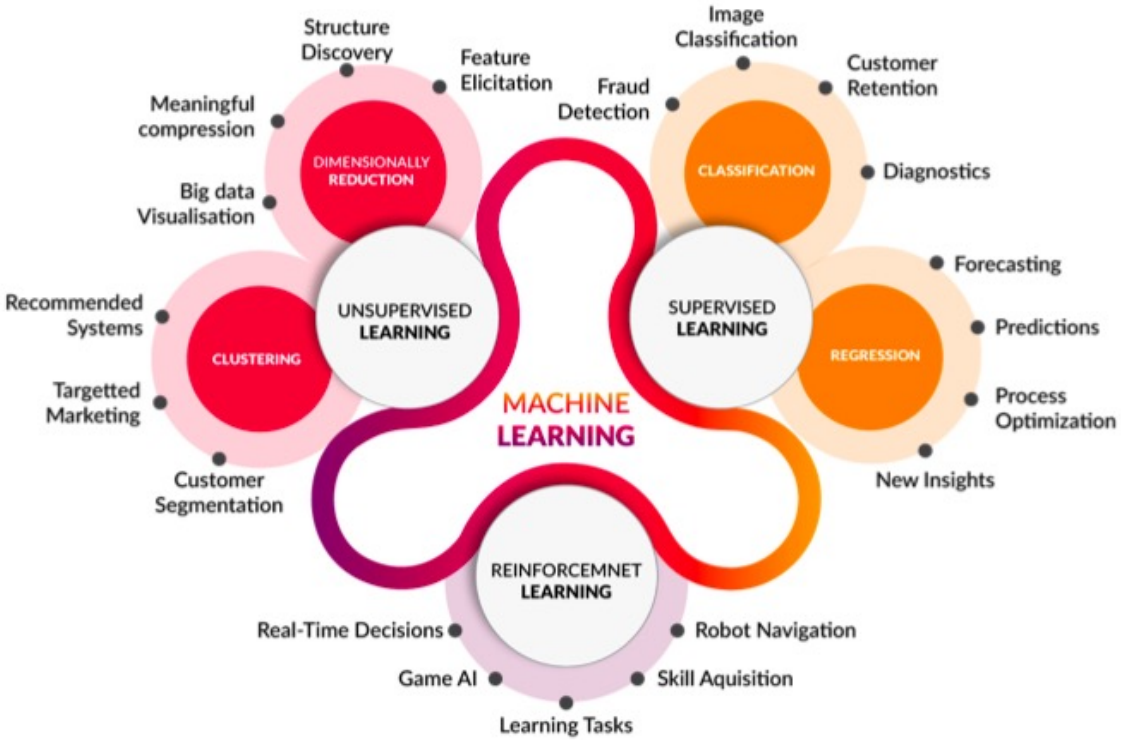
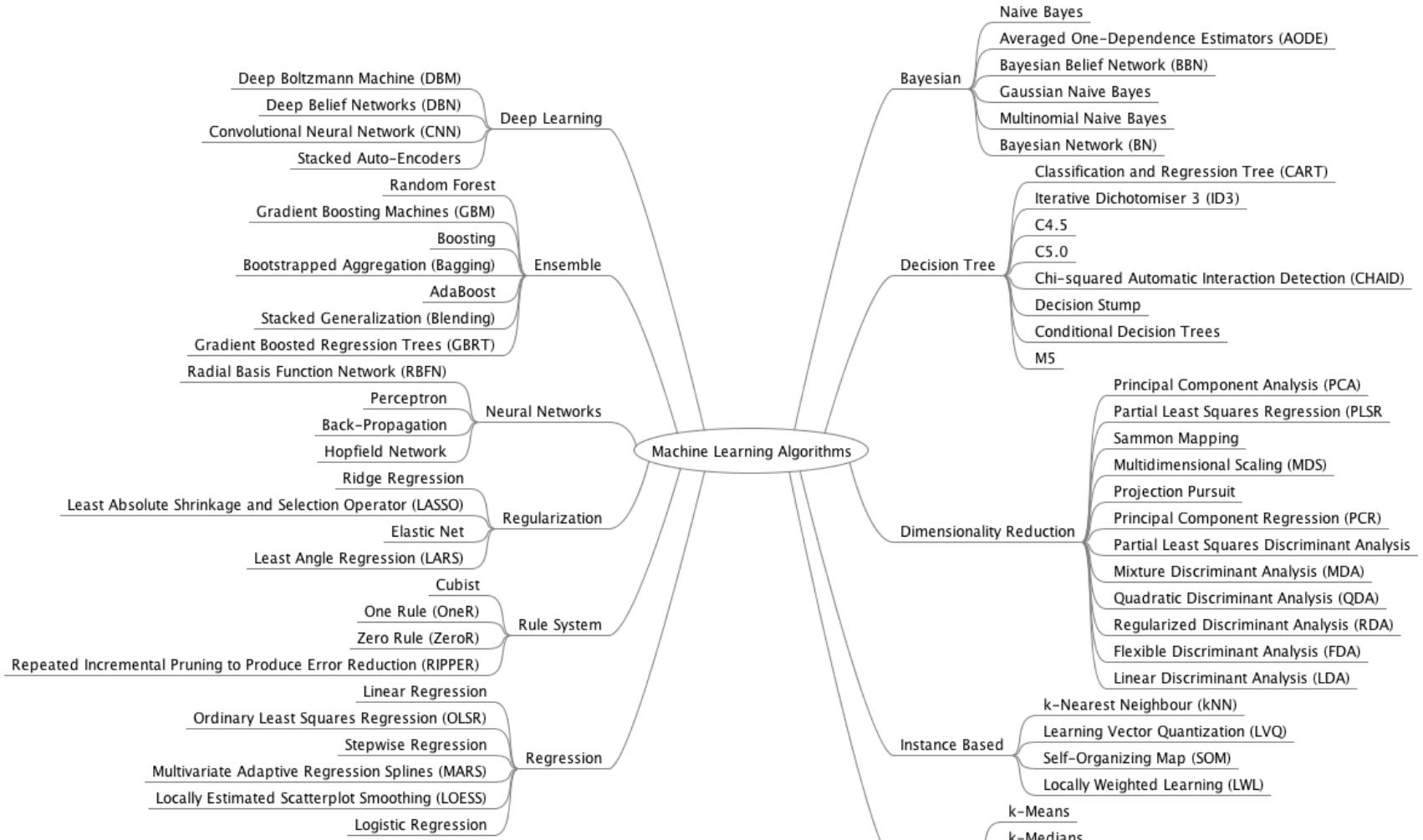


Figure: Machine Learning Map

(source: http://www.cognub.com/index.php/cognitive-platform/)



Source: https://mlforall.wordpress.com/2019/09/24/machine-learning-algorithms/

The Performance Measure (P)

- Quantitative measure to **evaluate** the abilities of machine learning to carry out the task *T*.
- Most common **performance metrics**:
  - accuracy,
  - error rate,
  - precision,
  - recall,
  - F1
- The task performance should be evaluated on a **test set – data** that is separate from the data used for training the machine learning system.

The Experience E

Machine learning should be getting better when facing more cases/experiences (data).

Supervised Learning

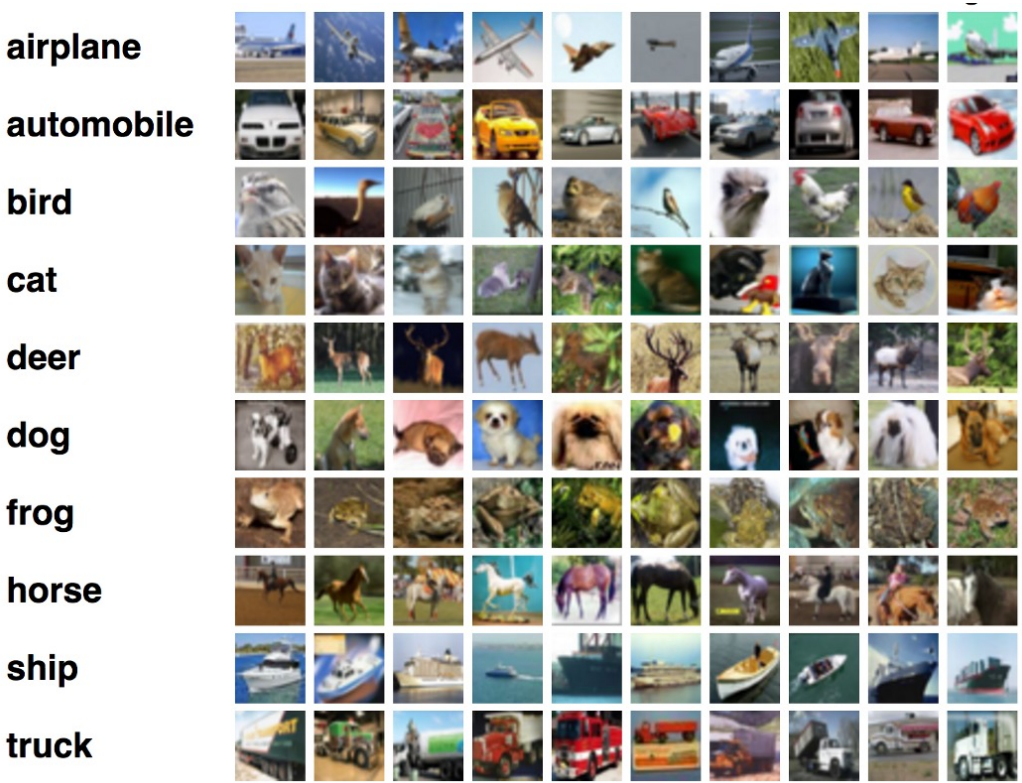


Figure: Classification Illustration
(source: towardsdatascience.com)

Classification

CAT

Data: (x,y)
x is features data, y is label/target

Goal: learn a *function* to map x → y

Regression vs Classification

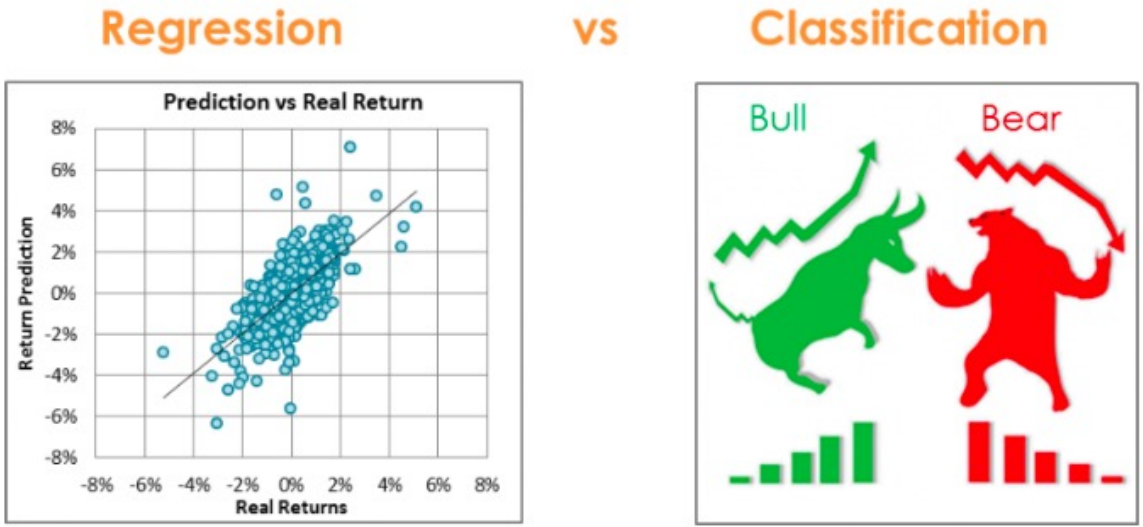


Figure: Regression vs Classification
(source https://quantdare.com/machine-learning-a-brief-breakdown/)

Supervised Learning workflow

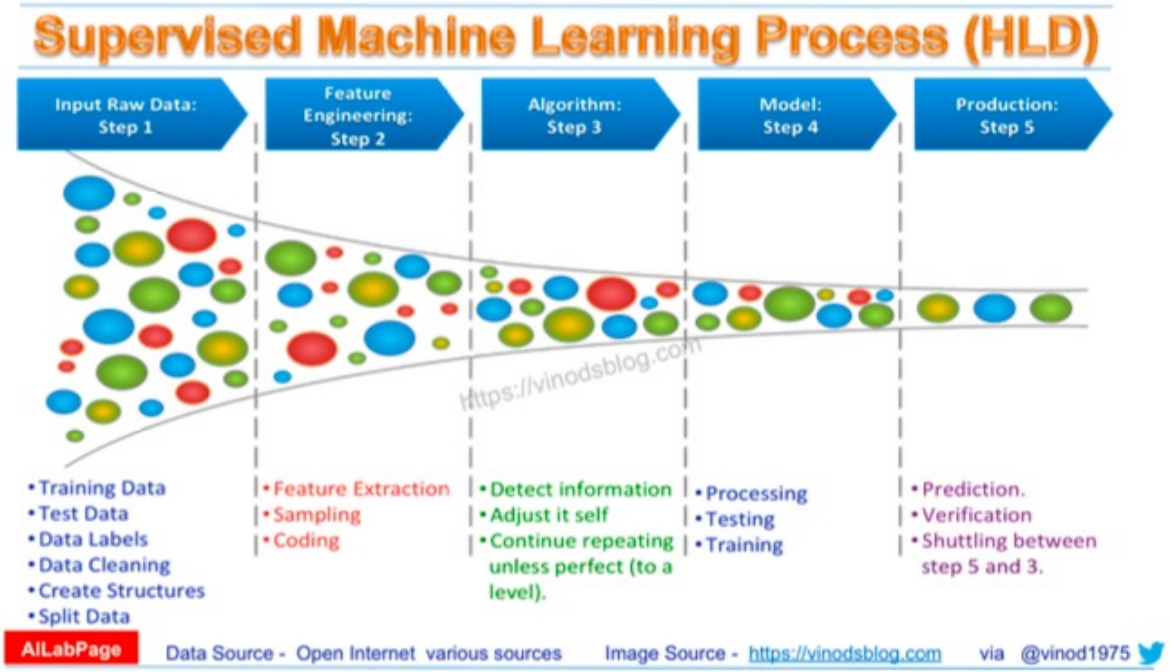


Figure: Supervised Learning workflow

(source: https://vinodsblog.com/2018/04/02/supervised-machine-learning-insider-scoop- for-labeled-data/)

Unsupervised Learning

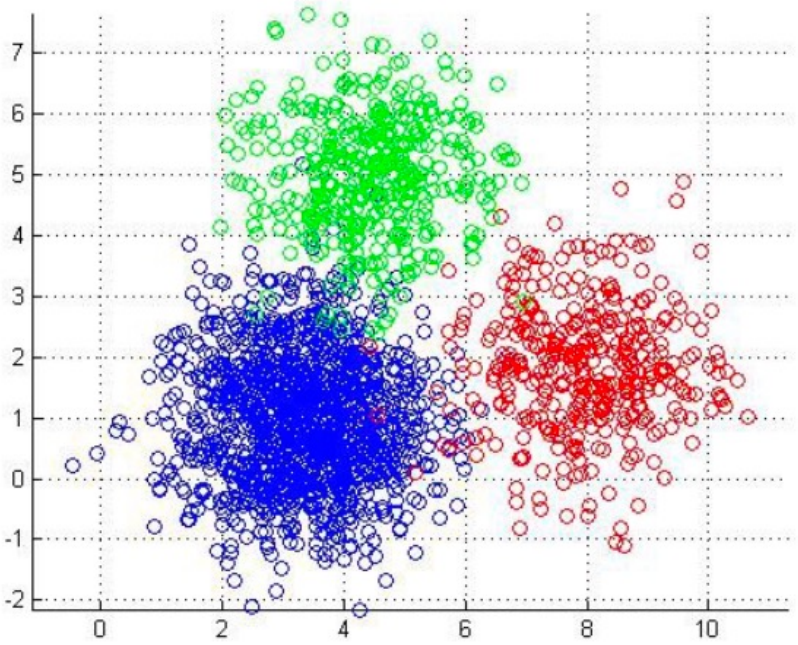


Figure: Unsupervised Learning

Data: x
x is features data, no labels!

Goal: learn some underlying hidden *structures or patterns* of the data

Supervised vs Unsupervised Learning

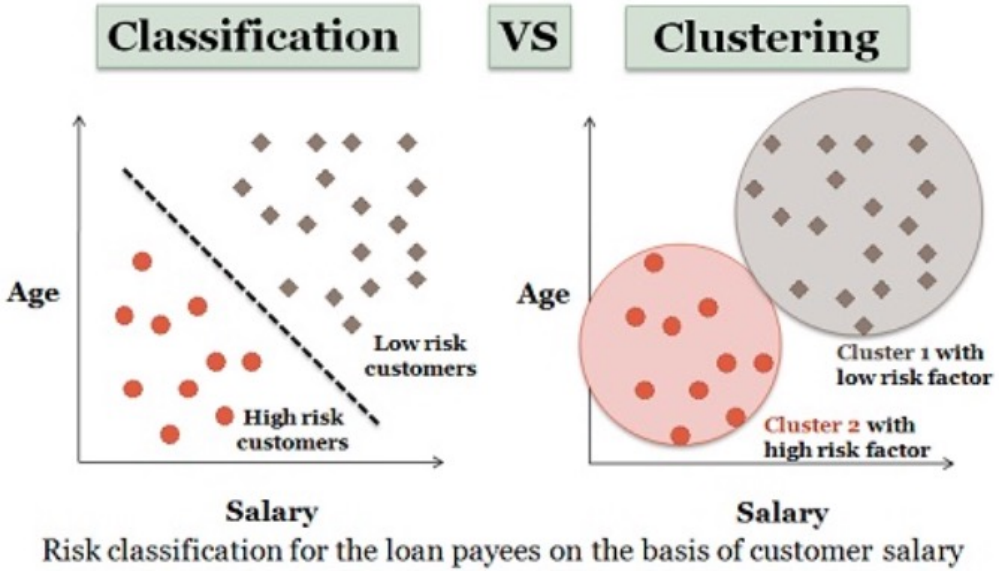


Figure: Unsupervised vs Supervised Learning

(source: https://techdifferences.com/difference-between-classification-and- clustering.html)

Unsupervised Learning Workflow

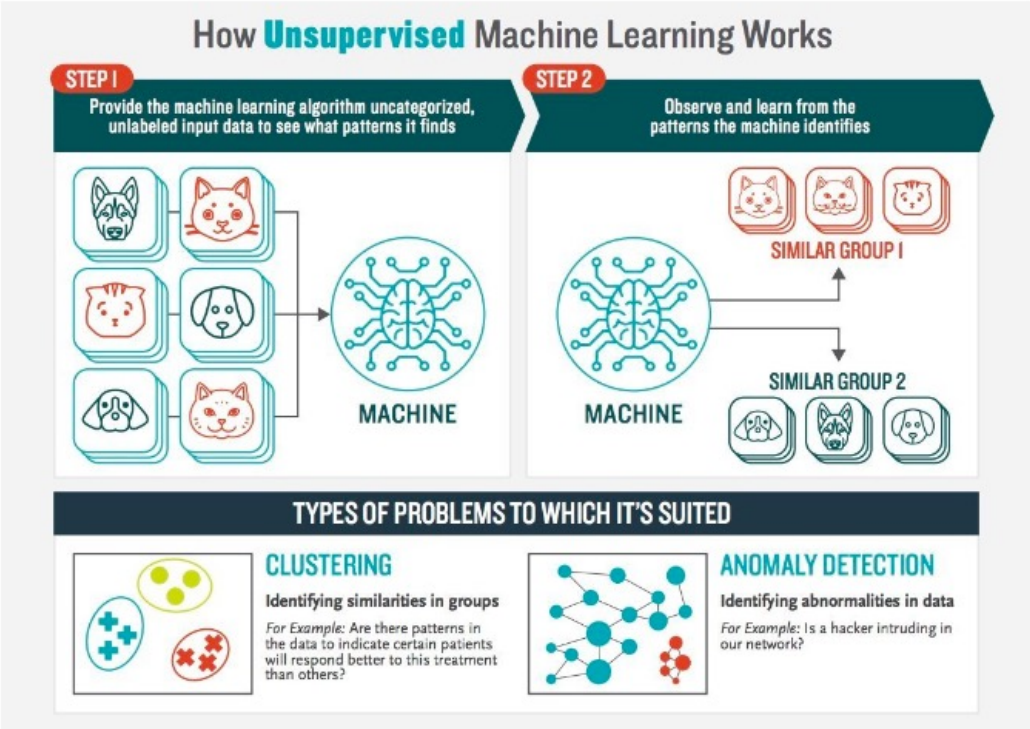


Figure: Unsupervised Learning Workflow
(source: boozallen.com)

Semi-supervised Learning

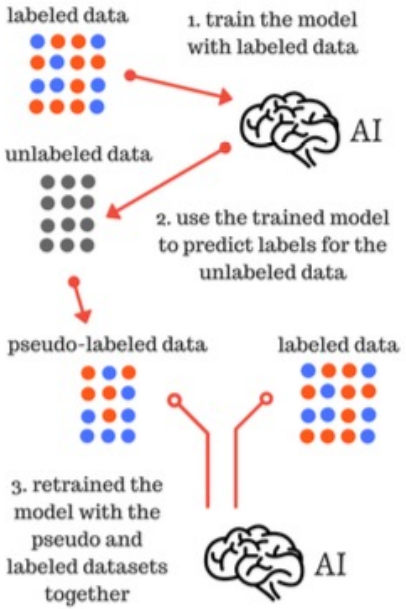
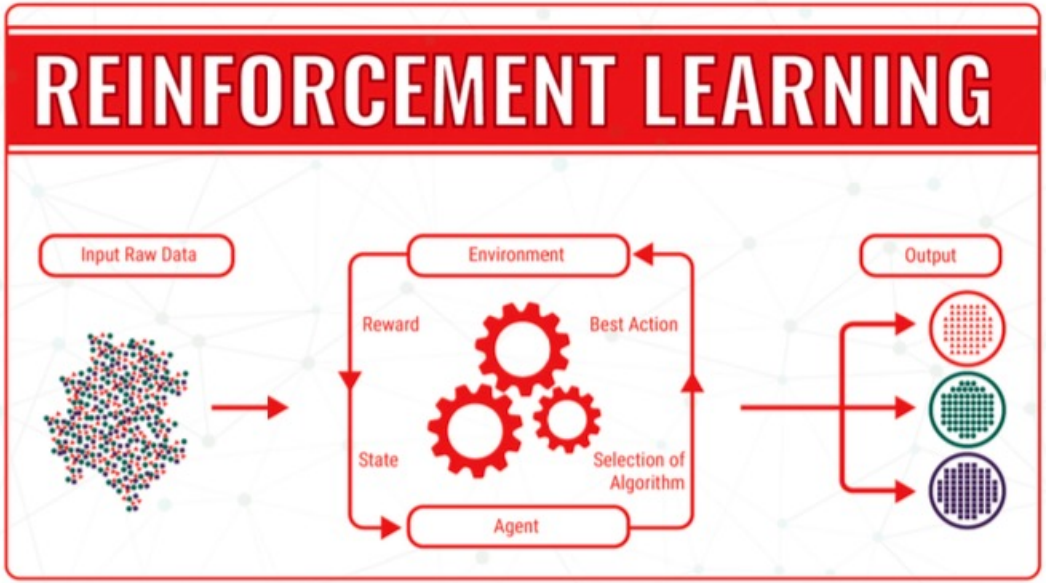


Figure: Semi-supervised Learning

(source: https://datawhatnow.com/pseudo-labeling-semi-supervised-learning/)

Reinforcement Learning



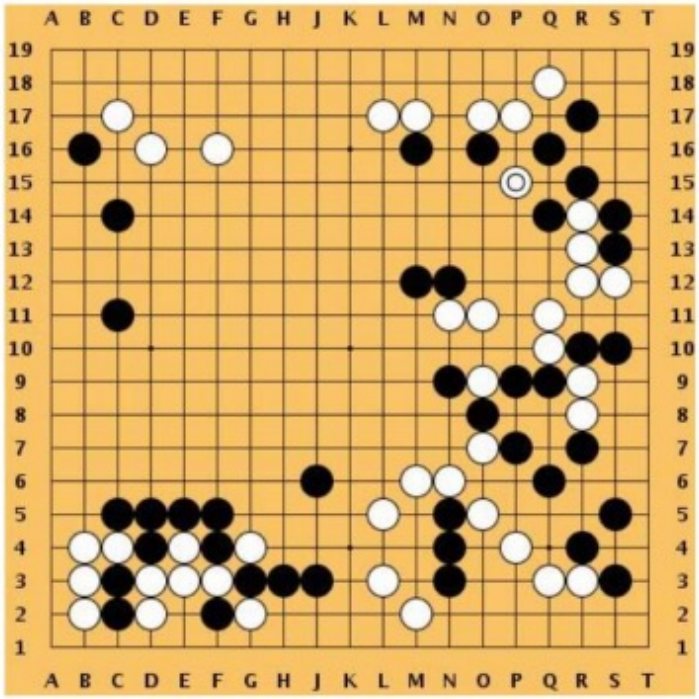
Problems involve an **agent** interacting with an **environment,** which provides numeric **reward (or penalty)** signals.

Goal: learn how to take action in order to maximize reward.

Figure: Reinforcement Learning

(source: http://www.sra.vjti.info/blog/machine- learning/introduction-to-reinforcement-learning-in-2-minutes/)

Reinforcement Learning



Go Game

Objective: Win the game!
State: Position of all pieces
Action: Where to put the next piece down
Reward: 1 if win at the end of the game, 0 otherwise

How does reinforcement learning work?

- Developers design a method of **rewarding desired behaviours** and **punishing negative behaviours**.
- This method **assigns positive values to the desired actions** to encourage the agent and **negative values to undesired behaviours**.
- This method programs the agent to **seek long-term** and **maximum overall rewards** to achieve an **optimal** solution.
- These **long-term goals** help prevent the agent from stalling on lesser goals.
- With time, **the agent learns** to avoid the negative and seek the positive.
- This learning method has been adopted in AI as a way of directing **unsupervised ML through rewards and penalties**.

Model Selection and Assessment

- **Model Selection**: Estimating performances of different models to choose the best one (produces the minimum of the test error)
- **Model Assessment**: Having chosen a model, estimate the prediction error on new data.

Splitting the Data

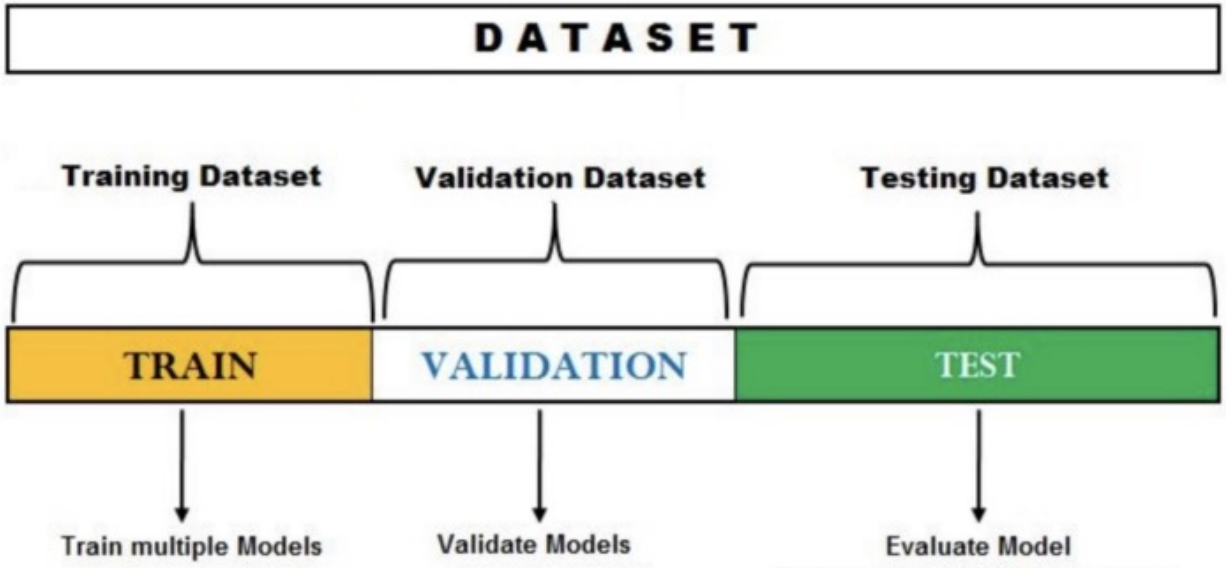


Figure: Data splitting
(source: https://medium.com)

Cross Validation

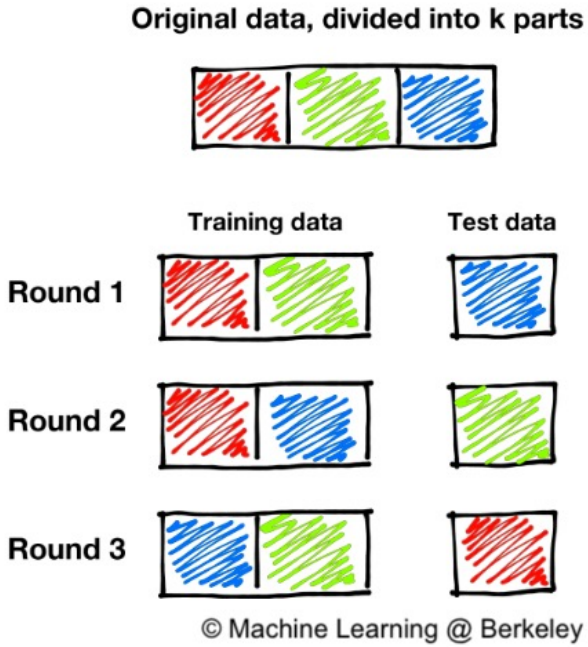


Figure: Cross Validation
(source: https://ml.berkeley.edu/blog/2017/07/13/tutorial-4/)

- **Cross-validation (CV)** is one of the techniques used to **test the effectiveness of ML models**, it is also a re-sampling procedure used to evaluate a model if we have **limited data**.
- **Goal**: test the model's ability to predict new data that was not used in estimating it, in order to **handle problems** like overfitting and selection bias, and to give an insight into an independent dataset (i.e., **how the model will generalize** an unknown dataset, for instance from a real problem).

THANK YOU