

An introduction to Machine Learning

A (very) gentle Introduction

Massimiliano Ruocco, PhD

*Senior Research Scientist, AAI Group, Telenor Research
Adjunct Associate Professor, DART Group, NTNU*



Outline

— — —

- Introduction
 - Supervised ML
 - Unsupervised ML
 - Reinforcement Learning
 - Deep Learning
 - Conclusion
-
- Exercise!

Introduction

What is Machine Learning

— — —

What is Machine Learning

— — —

Machine learning is an **application** of artificial intelligence (AI) that provides **systems** the ability to **automatically learn** and improve from **experience without** being explicitly **programmed**.

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed

Machine learning is an artificial intelligence (AI) discipline geared toward the technological development of human knowledge. Machine learning **allows computers** to handle **new situations** via analysis, self-training, observation and **experience**.

Taxonomy of Machine Learning

— — —

Artificial Intelligence



Machine Learning

Taxonomy of Machine Learning

— — —

Artificial Intelligence

Machine Learning

Supervised Learning

- Classification
- Regression

Unsupervised Learning

- Clustering

Reinforcement Learning

Why Machine Learning

— — —

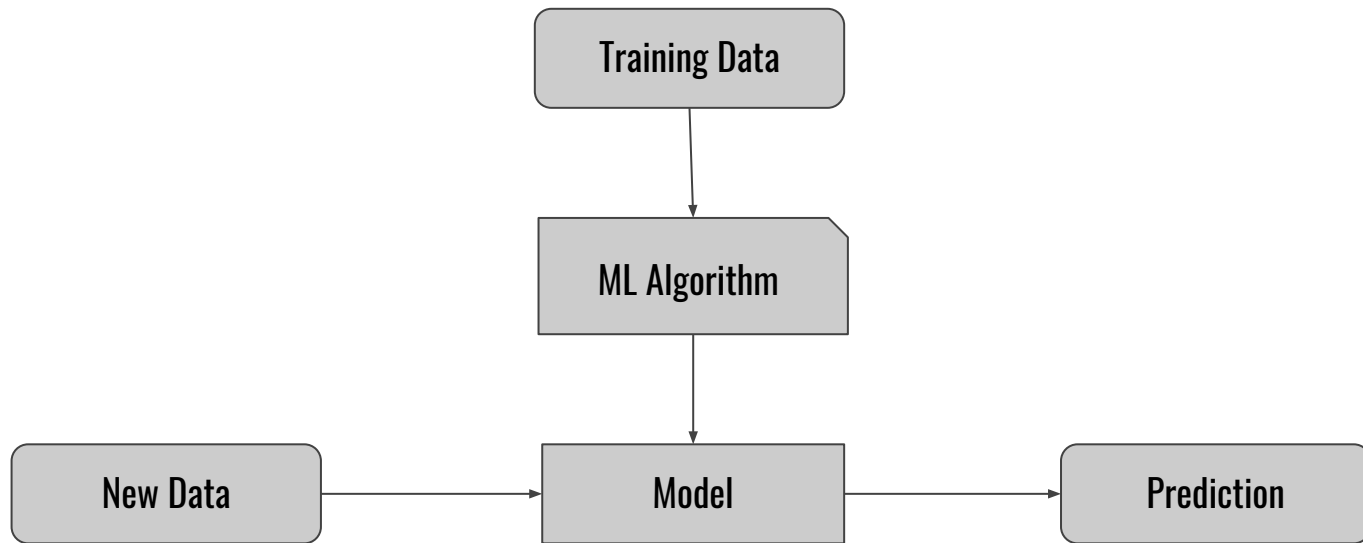
Machine Learning Examples

Retail	Marketing	Healthcare	Telco	Finance
<ul style="list-style-type: none">• Demand forecasting• Supply chain optimization• Pricing optimization• Market segmentation and targeting• Recommendations	<ul style="list-style-type: none">• Recommendation engines & targeting• Customer 360• Click-stream analysis• Social media analysis• Ad optimization	<ul style="list-style-type: none">• Predicting Patient Disease Risk• Diagnostics and Alerts• Fraud	<ul style="list-style-type: none">• Customer churn• System log analysis• Anomaly detection• Preventative maintenance• Smart meter analysis	<ul style="list-style-type: none">• Risk Analytics• Customer 360• Fraud• Credit scoring

Supervised ML

Introduction

— — —



Classification vs Regression

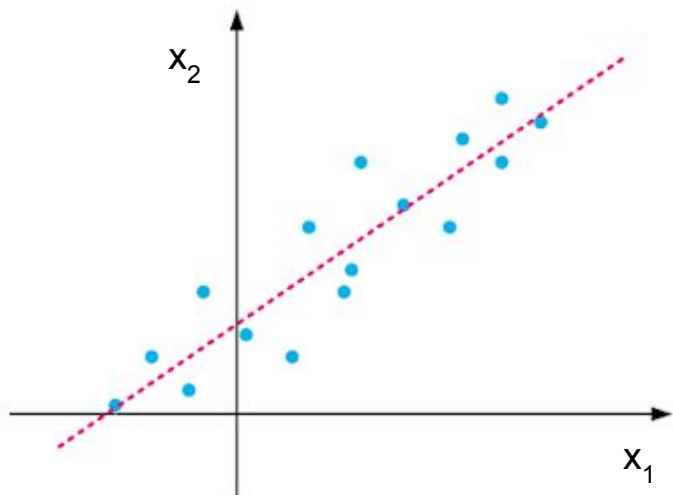
Regression: predict a continuous numerical value.

How much will be the price of this stock?

Classification: assign a label to a given observation.

Is this a picture of a car or a motorcycle?

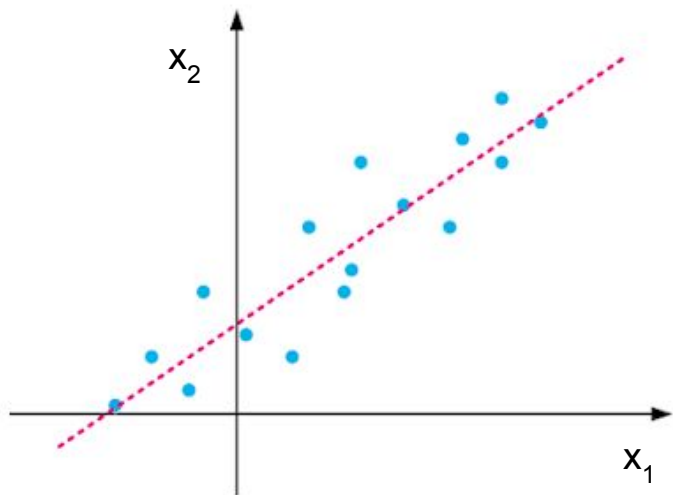
Classification vs Regression



$$x_2 = f(x_1, \beta)$$

Classification vs Regression

— — —

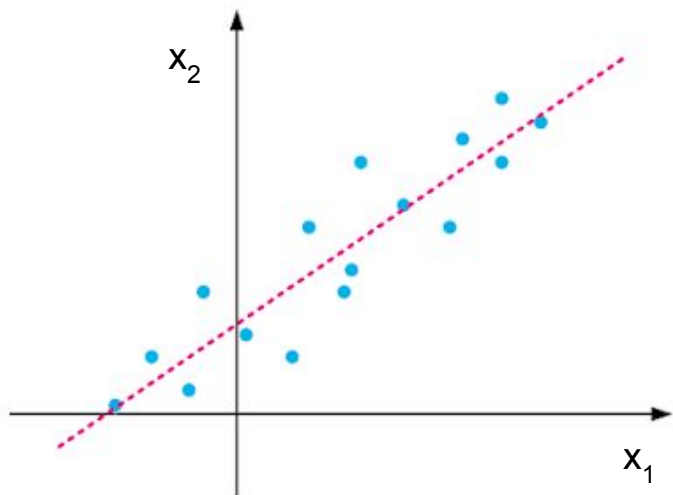


$$X_2 = f(X_1, \beta)$$

Diagram illustrating the regression function $X_2 = f(X_1, \beta)$ with labels and arrows:

- X_2 is labeled as the **Dependent variable** (indicated by an upward arrow).
- X_1 is labeled as the **Independent variable** (indicated by an upward arrow).
- f is labeled as the **Regression function** (indicated by a downward arrow).
- β is labeled as the **Parameters of the function f** (indicated by a downward arrow).

Classification vs Regression



$$X_2 = f(X_1, \beta)$$

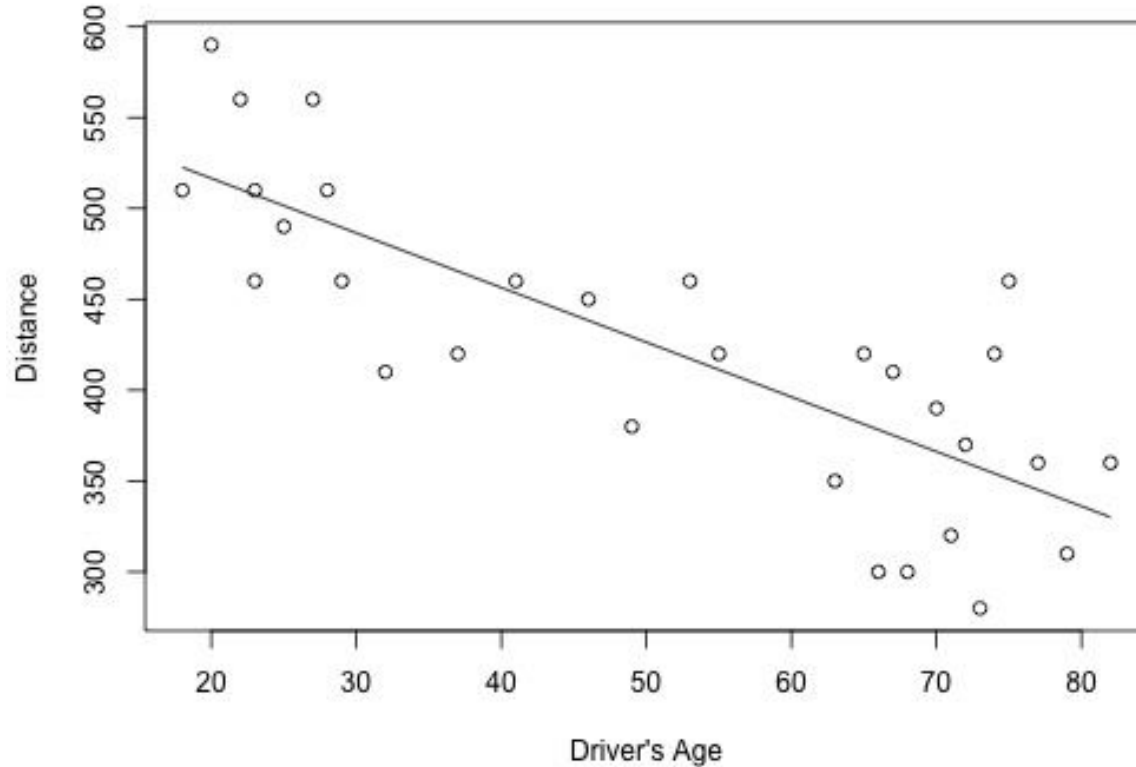
Diagram illustrating the regression function f and its parameters β .

The equation $X_2 = f(X_1, \beta)$ is shown, where:

- X_2 is the **Dependent variable**.
- X_1 is the **Independent variable**.
- f is the **Regression function**.
- β represents the **Parameters of the function f** .

Main Task: Estimating f , giving a set of training data \mathbf{X}

Classification vs Regression: *Example*



Classification vs Regression

— — —

Regression Algorithms

- Logistic Regression
- Linear Regression
- Support Vector Regression
- Regression Trees
- Least Angle Regression
- ...

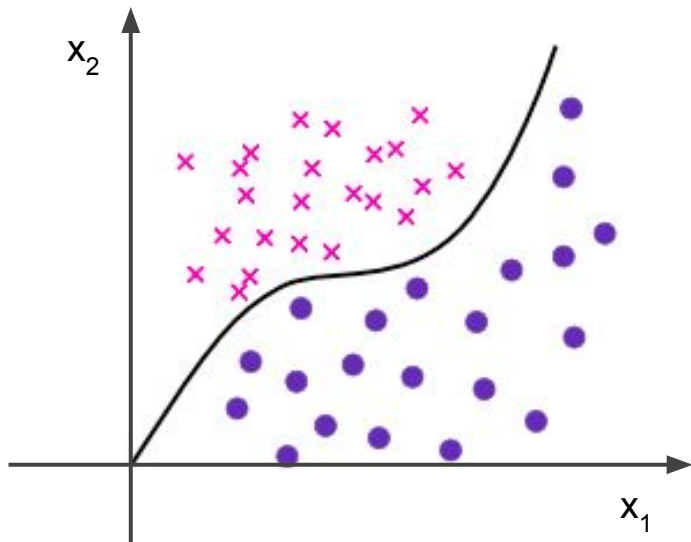
Classification vs Regression: *Linear Regression*

— — —

x_1

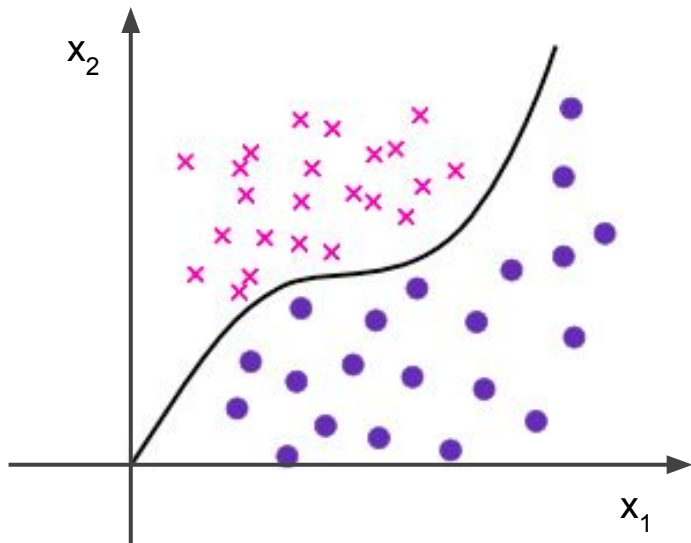
Classification vs Regression

— — —



$$Y = f(\mathbf{X})$$

Classification vs Regression

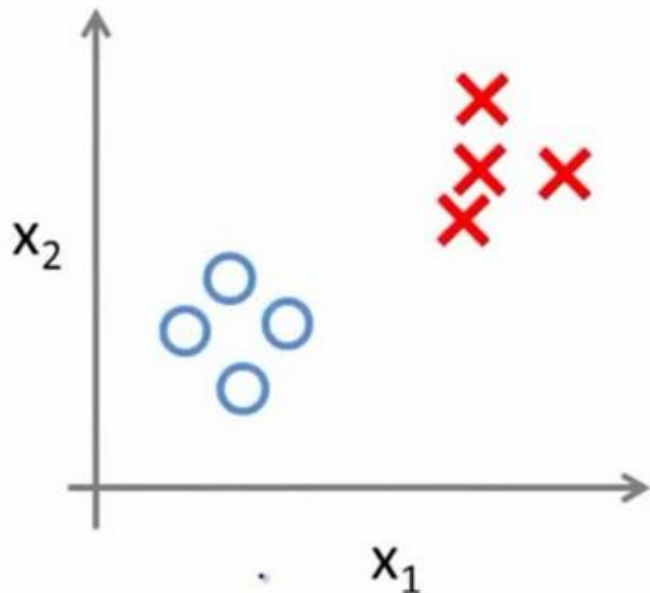


Main Task: Estimating f , giving a set of training data \mathbf{X}

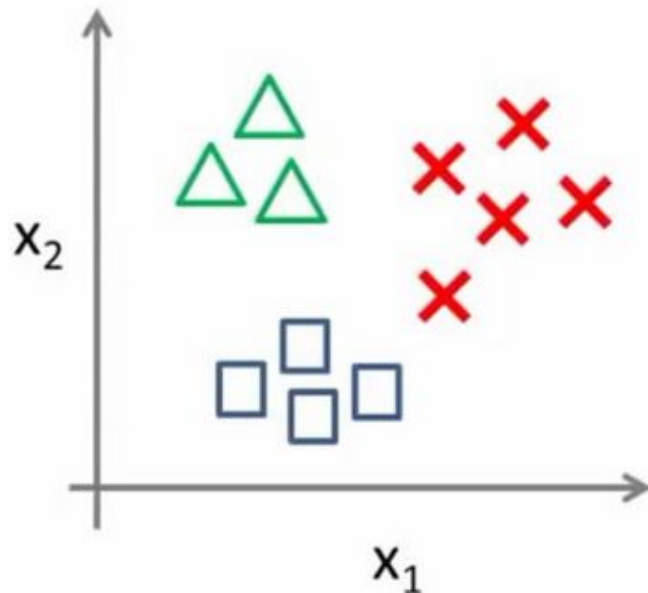
$$f: \mathbb{R}^d \rightarrow \{L_1, \dots, L_N\}$$
$$\begin{array}{ccc} & \downarrow & \\ Y & = & f(\mathbf{X}) \\ \uparrow & & \uparrow \\ \text{Label} & & \text{Input} \end{array}$$

Classification vs Regression

Binary Classification



Multiclass Classification



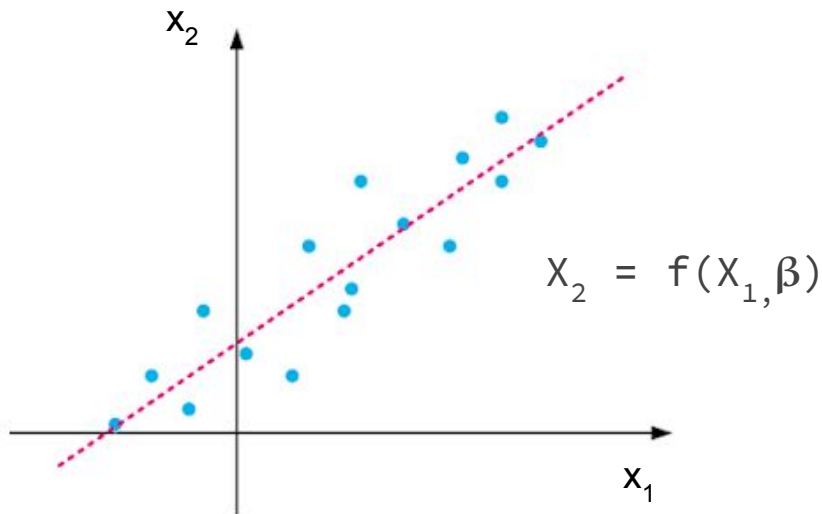
Classification vs Regression

— — —

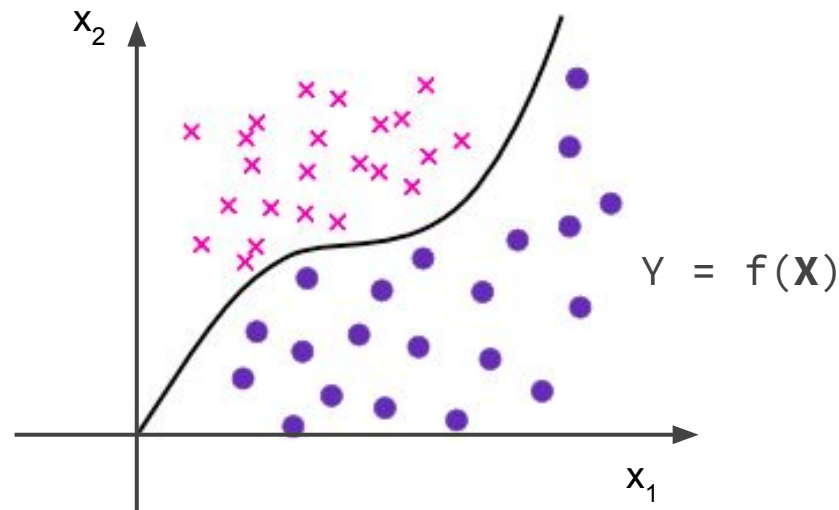
Classification Algorithms

- Support Vector Machines (SVM)
- Neural Networks
- Naive Bayes Classifier
- Nearest Neighbors (kNN)
- ...

Classification vs Regression: Summary



Input - Continuous
Model - Best fit Line
Evaluation - Sum of squared error
Inference - Predict the value of x_2 given the value of x_1



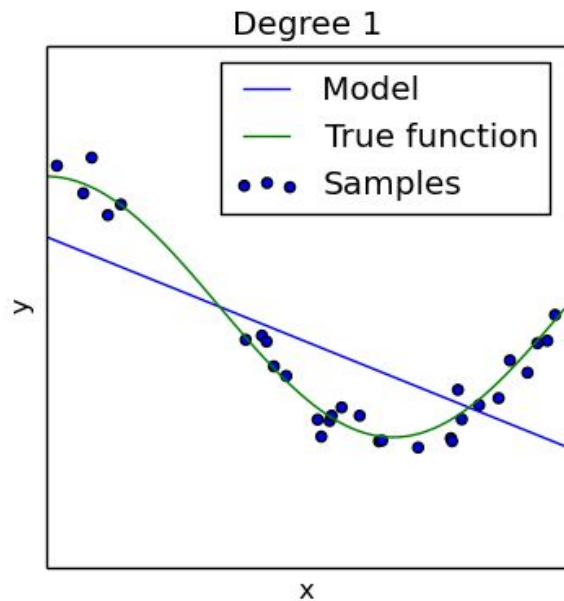
Input - Discrete
Model - Decision Boundary
Evaluation - Accuracy
Inference - Predict the label of a given \mathbf{X}

Overfitting and Underfitting

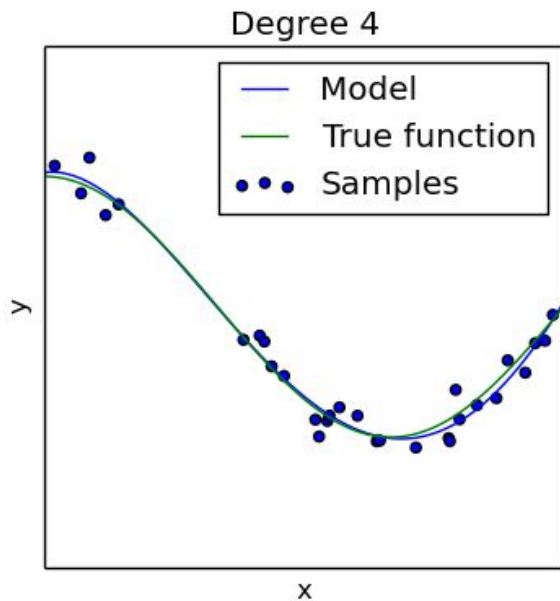
— — —

Overfitting and Underfitting

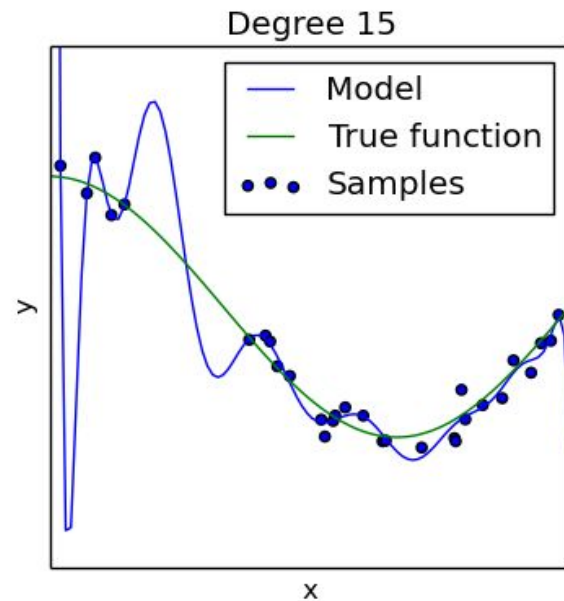
— — —



Underfitting



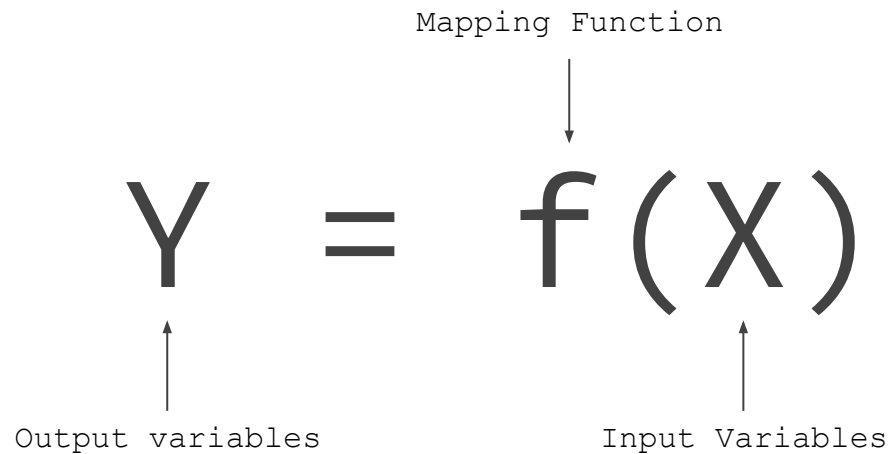
Optimal



Overfitting

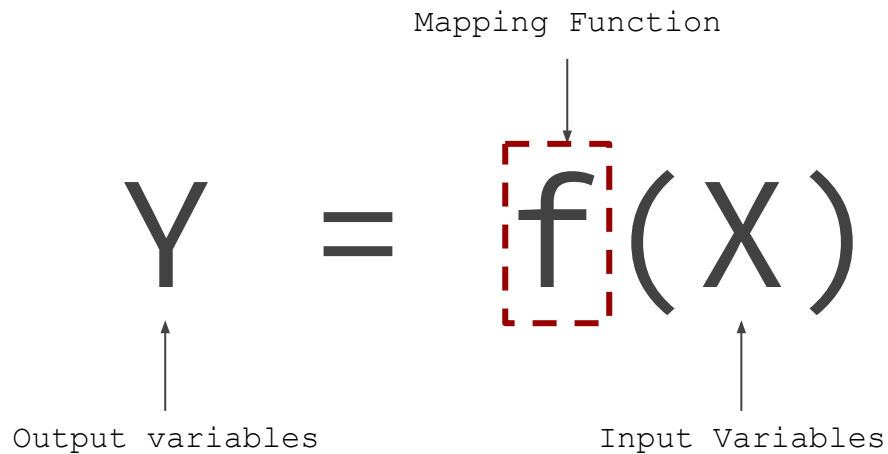
Parametric vs Nonparametric Methods

— — —



Parametric vs Nonparametric Methods

— — —



Which kind of function is better is better to use for approximating f ?

Parametric vs Nonparametric Methods

— — —

- **Assumption** on the underlying **function**
- **Fixed** Number of Parameters to learn
- I.e.: *Linear Regression*

$$x_2 = \mathbf{b}x_1 + \mathbf{a}$$

Example

- Logistic Regression
- Linear Discriminant Analysis
- Perceptron
- Naive Bayes
- Simple Neural Network
- Linear SVM

Pros

- Faster
- Simpler
- Less data
- Easy on memory

Cons

- Constrained
- Limited Complexity
- Poor Fit

Parametric vs Nonparametric Methods

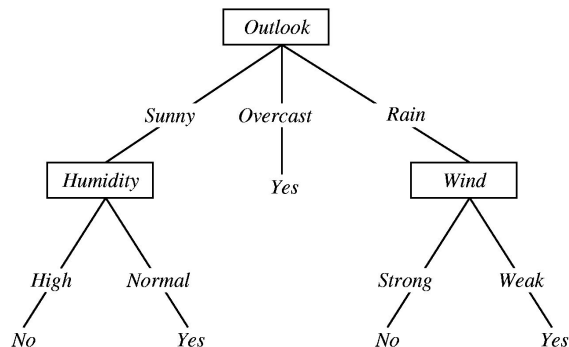
- **No Assumption** on the underlying mapping **function**

Pros

- Flexible
- Better Performance

Example

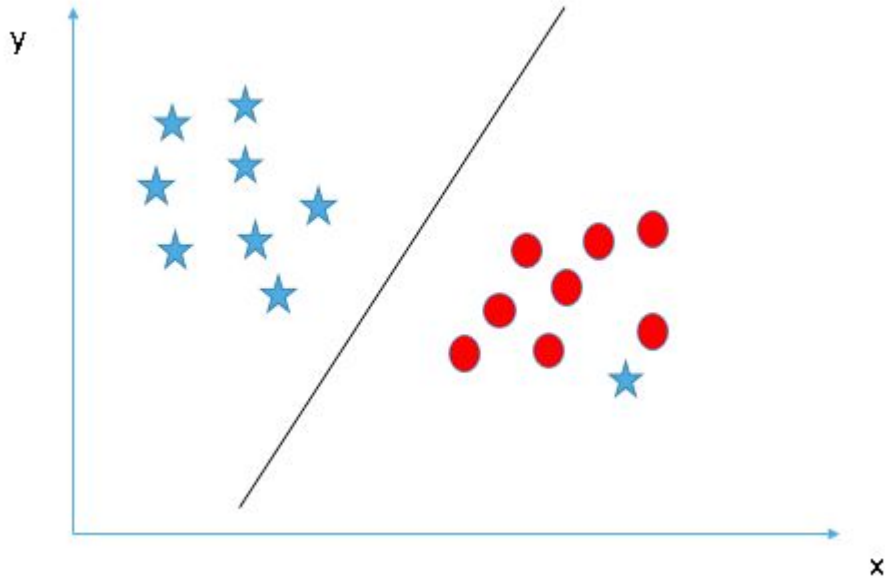
- k-NN
- Decision Trees
- RBF-Kernel SVM



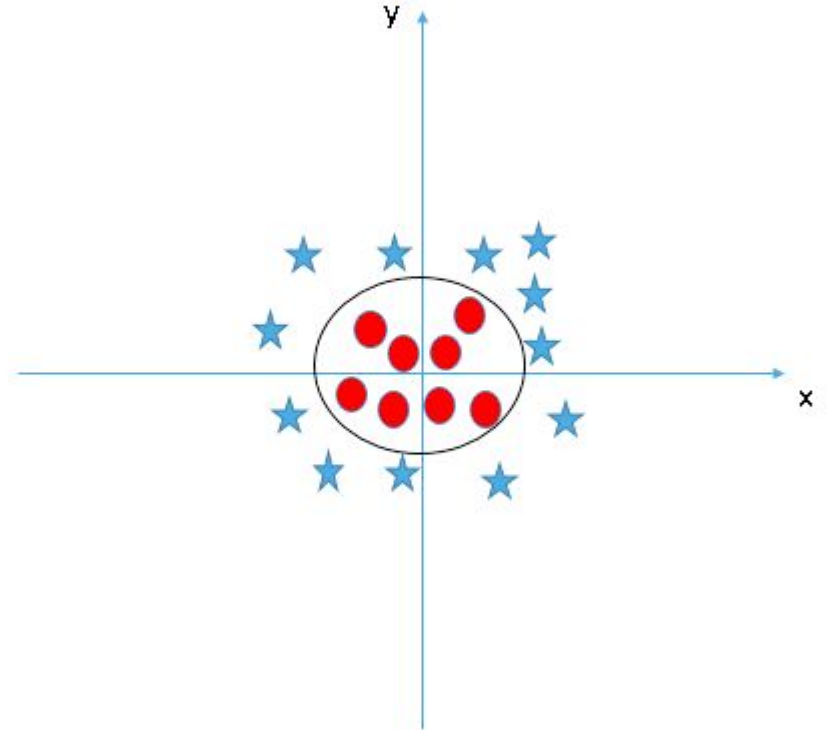
Cons

- More Data
- Slower
- Overfitting

Linear vs Non-Linear



Linearly Separable Data



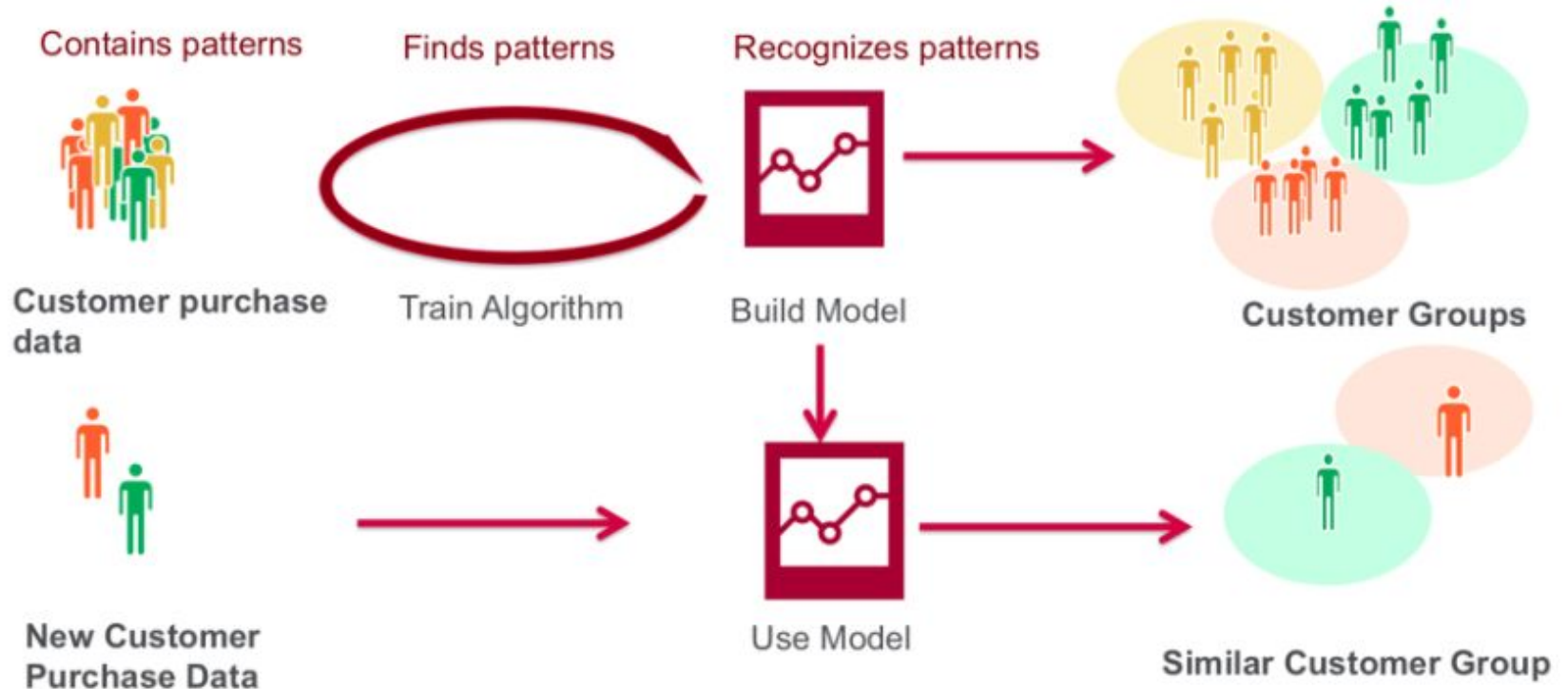
Non Linearly Separable Data

Unsupervised ML

Introduction

— — —

Introduction: *Example*

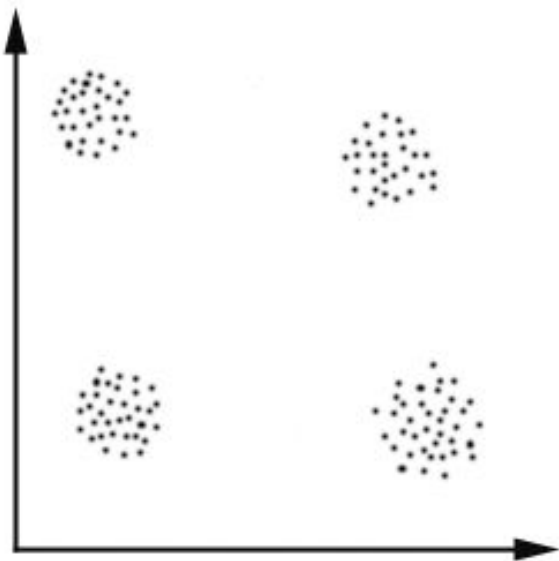


Clustering

— — —

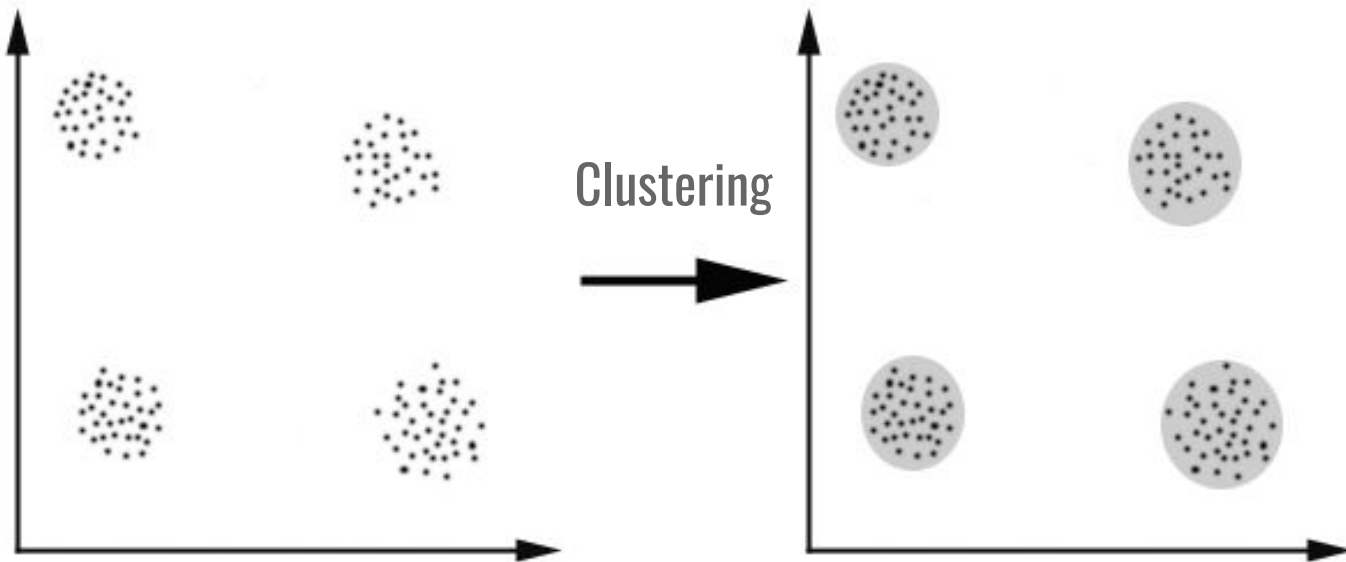
Clustering

— — —



Clustering

— — —



Clustering: *Applications*

— — —

- **Biology:** find similar entities/organisms
- **Information Retrieval:** Search results grouping.
- **Marketing:** Grouping similar customers.
- **Climate:** Find pattern of weather behaviour
- **Document/Text** categorization
- **Network security:** anomaly detection (finds what is not similar, the outliers from clusters).

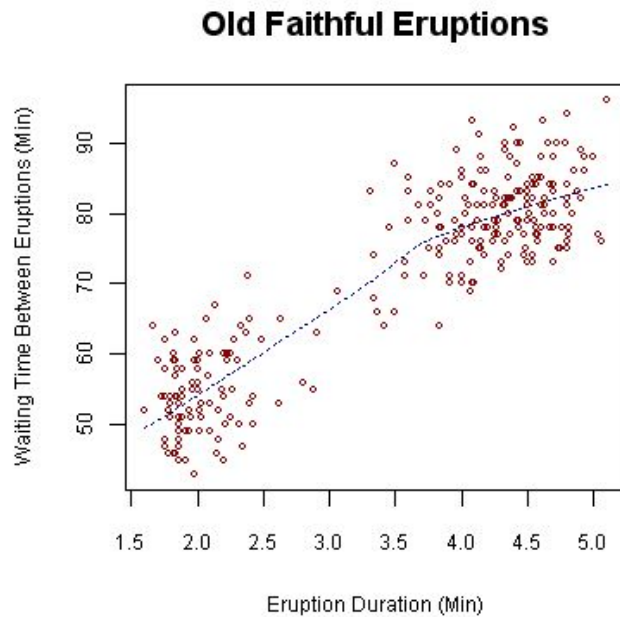
Clustering: *Methods*

— — —

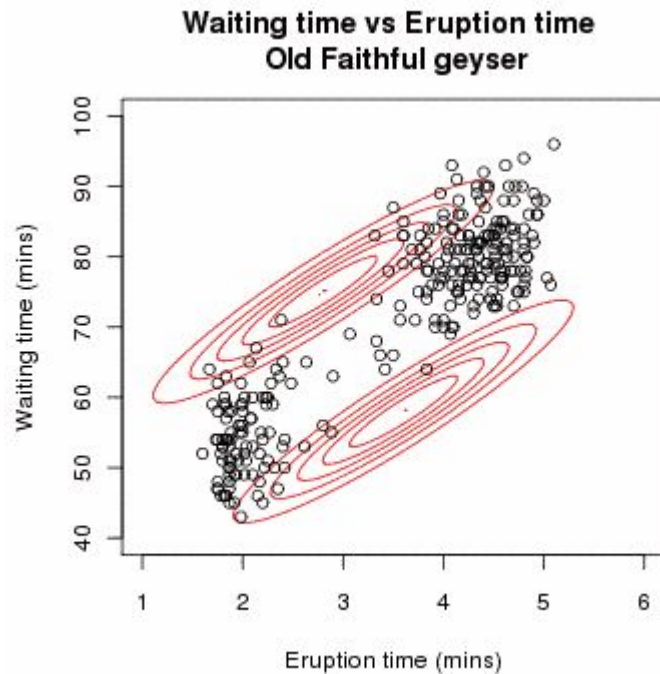
- **K-means Clustering**
- **Principal Component Analysis (PCA)**
- **Gaussian Mixture Model**
- **Self-Organizing Map (SOM)**
- **Hidden Markov Models (HMM)**
- **...**

Clustering - *GMM* demo

— — —



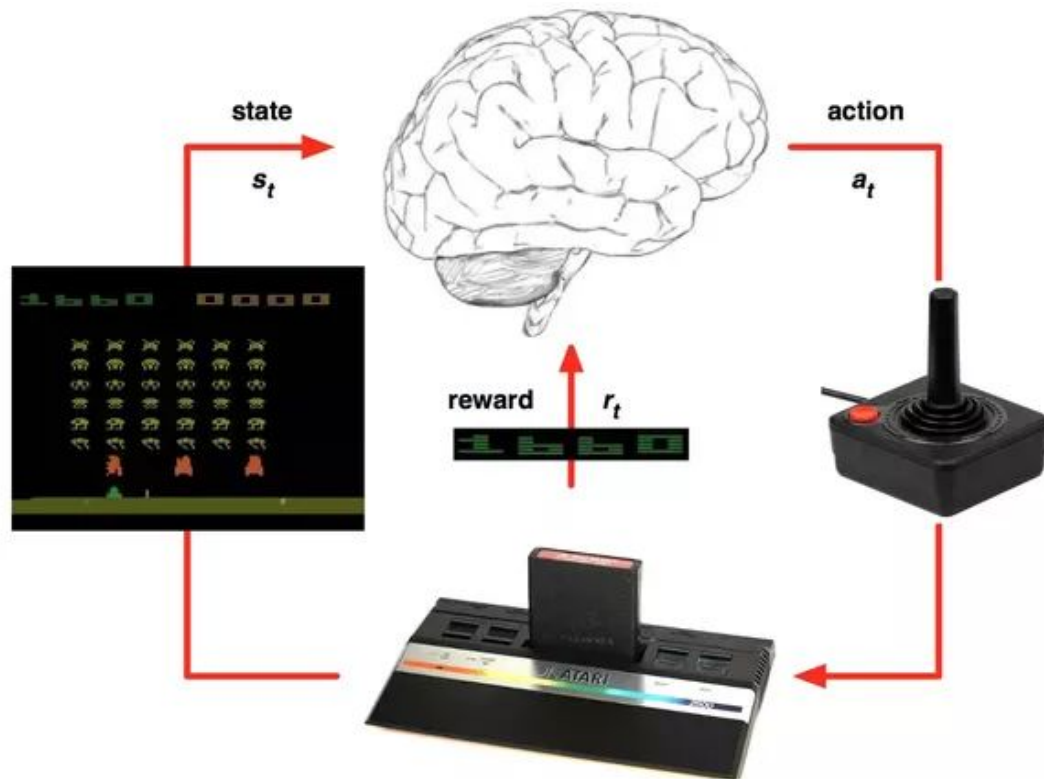
Clustering - *GMM* demo



Reinforcement Learning

Introduction

— — —



Application: *Facts on Alpha Go Zero*

— — —

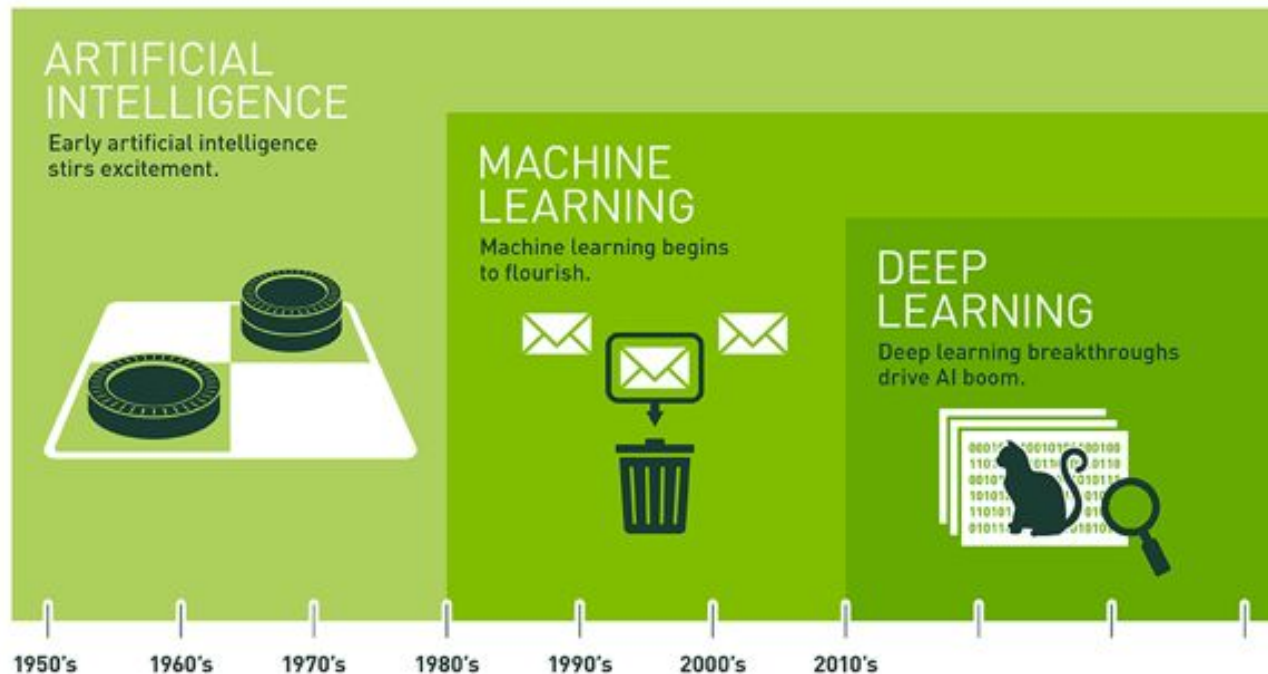
- Google DeepMind **AlphaGo Zero** beaten 100-0 the previous **AlphaGo**
- **AlphaGo** needed 100K go games to be trained
- **AlphaGo Zero** only programmed with the basic rules of Go
- **AlphaGo** beaten **18 times** the world champions **Lee Se-dol**
- **AlphaGo Zero** started beating (90%) the strongest AlphaGo after 40 days of training



Deep Learning

Introduction: *AI* vs *ML* vs *DL*

— — —

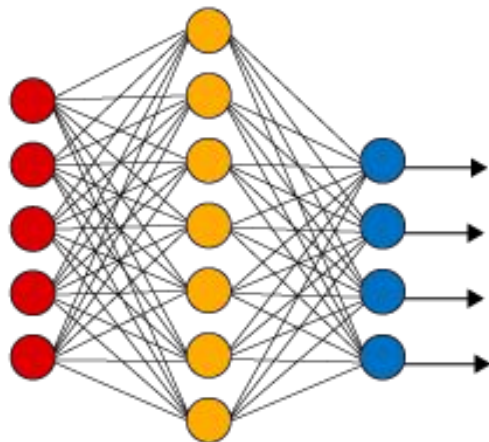


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Introduction: *Deep Network*

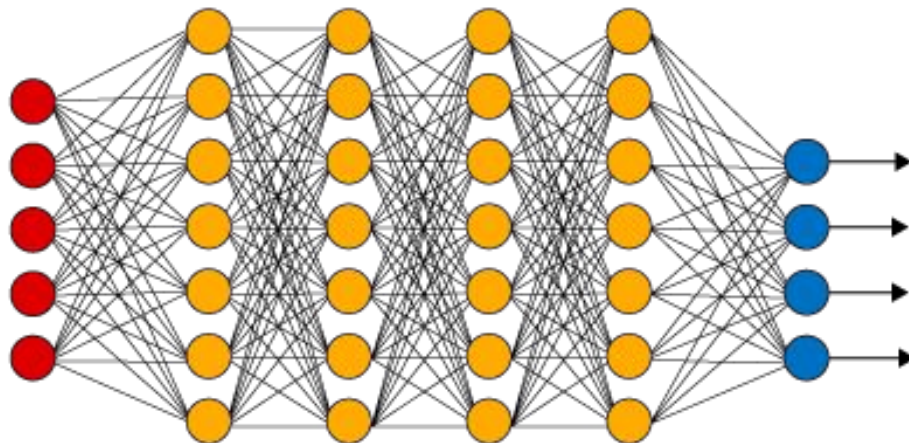
— — —

Simple Neural Network



● Input Layer

Deep Learning Neural Network

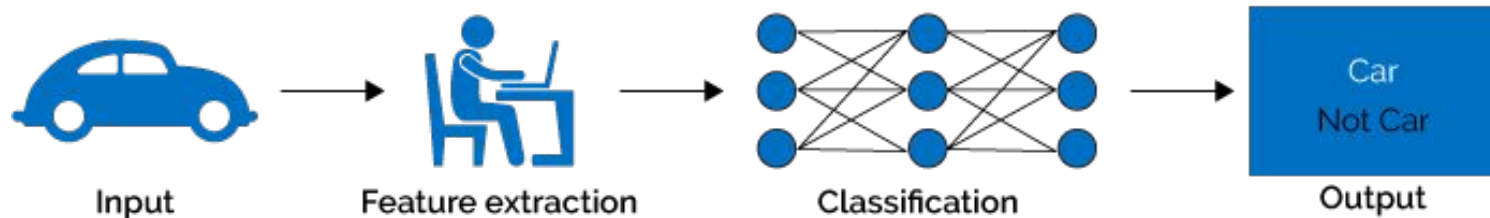


● Hidden Layer

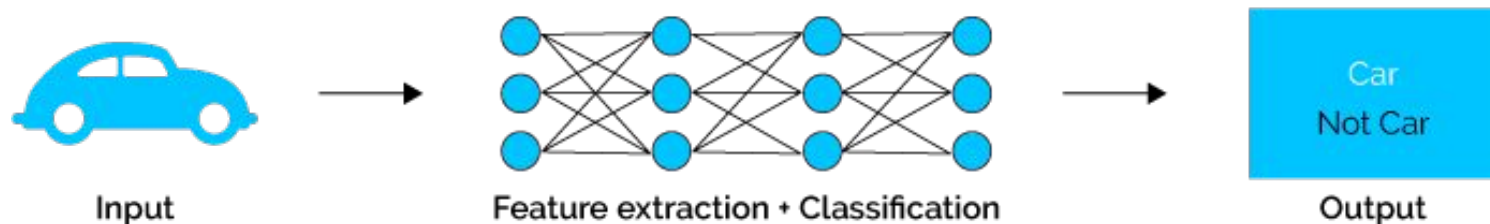
● Output Layer

Introduction

Machine Learning

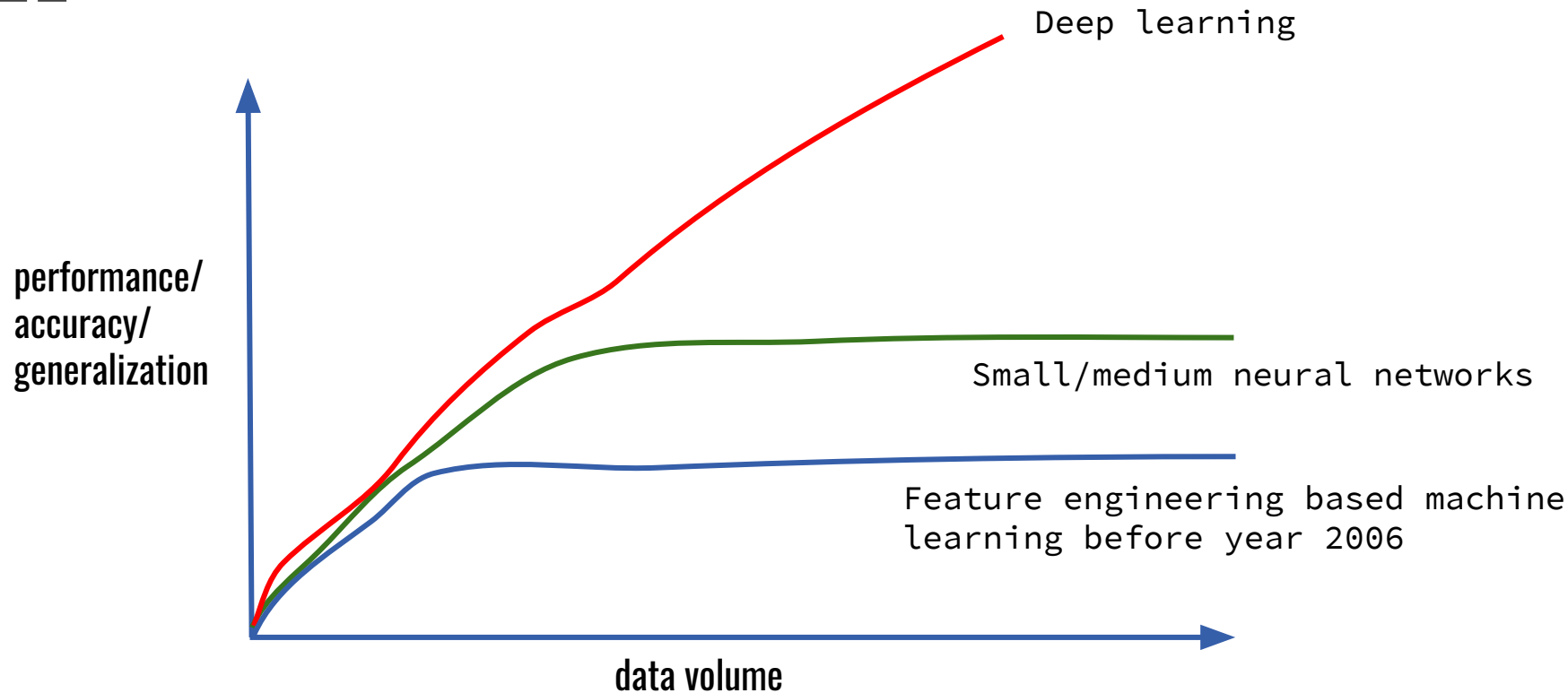


Deep Learning



Introduction: *Motivation on Data Volume*

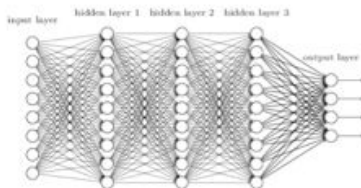
— — —



Introduction: *Main Architectures*

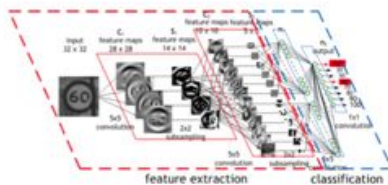
providing lift for
classification and
forecasting models

Deep
Neural
Networks



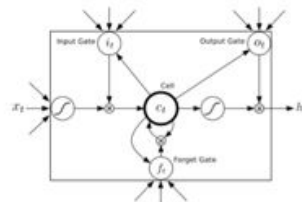
feature extraction
and classification of
images

Convolutional
Neural
Networks



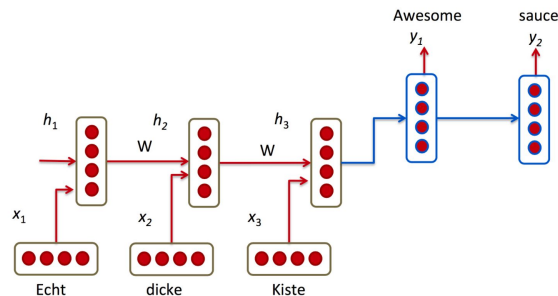
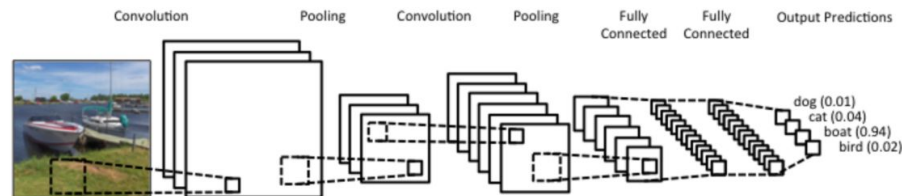
for sequence of events,
language models, time
series, etc.

Recurrent
Neural
Networks



RNN vs CNN

- **CNN:** Neural network able to recognize patterns across the space (i.e.: component of an image)
- **RNN:** NN able to capture pattern from sequential data
- CNN + RNN in joint architectures!



Recurrent Neural Network: Applications

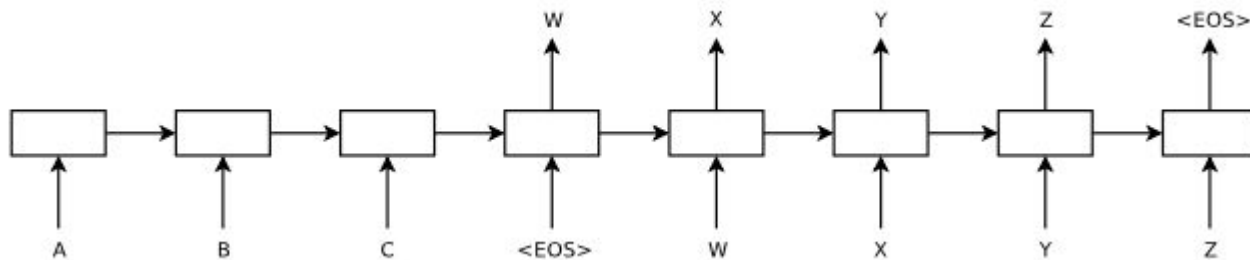
— — —

- Sentence Modelling
- Click Prediction
- Location Prediction
- Language Translation
- Sentiment Analysis
- Image Captioning and Description
- Speech Recognition
- Question/Answering Systems
- Text Generation

RNN Applications: Machine Translation

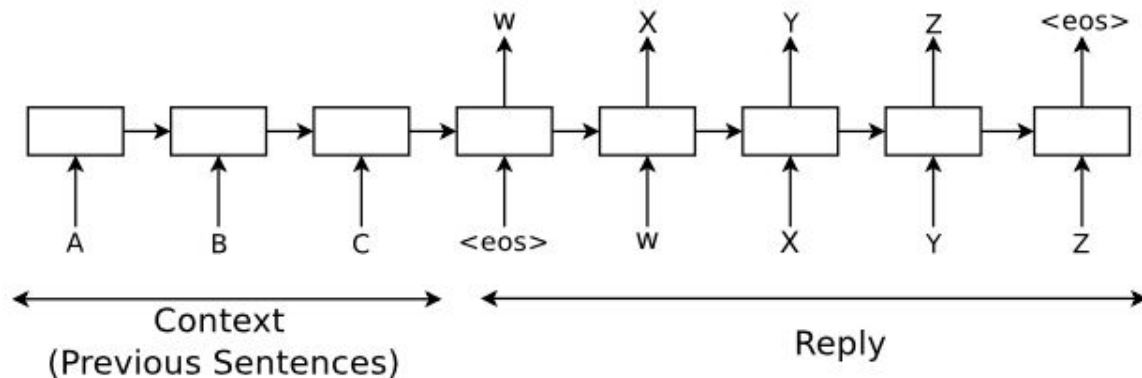
— — —

- Sequence to Sequence Model* [*Sutskever, Vinyals, Le, 2014*]
 - LSTM based model



RNN Applications: Conversational Systems

- — —
- Neural Conversational Model*
[Sutskever, Vinyals, Le, 2014]
 - LSTM based model



Machine: hi

Human: hello

Machine: could you please let me know what are the operating systems you are using ?

Human: i am using linux

Machine: ok it was in the ssl vpn on which you stated it was broken

Machine: please check if you are logged in to corp network

Machine: what is the error that you are running please

Human: i am seeing an error related to vpn

Machine: what is the error message that you are getting when connecting to vpn using network connect ?

Human: connection refused or something like that

Machine: may i know the version of network connect you connect ?

Human: i am not sure i know that

Machine: is the network connect prompting for any user certificate

Machine: ?

Human: yes

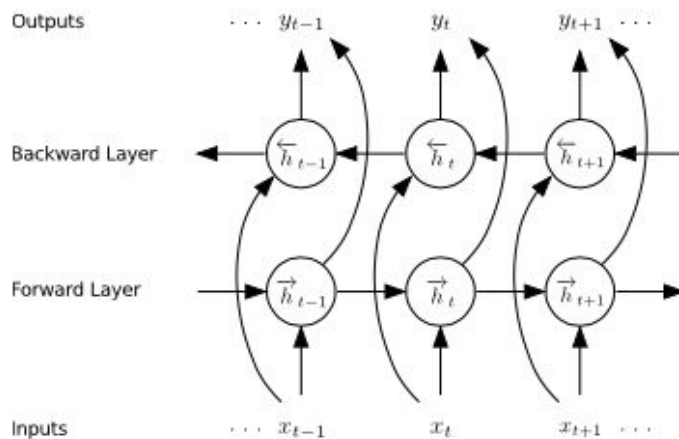
Machine: are you at home or at work at the moment ?

Machine: office ?

Human: i am at home

RNN Applications: Speech Recognition

- — —
- End to End Speech Recognition*
[Graves, Jaitly, 2014]
 - Bi-LSTM based model

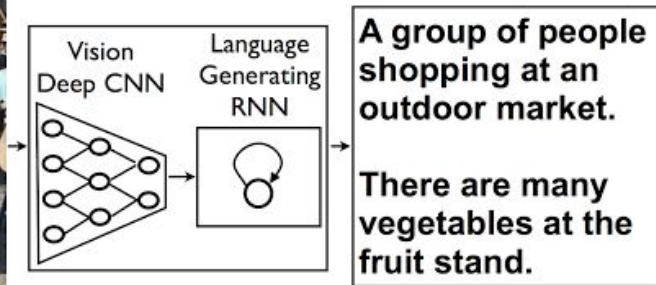


*Towards End-to-End Speech Recognition with Recurrent Neural Networks, A.Graves, N.Jaitly (2014)

RNN Applications: Image Captioning

— — —

- Generating Captions describing an Image
[O.Vinyals et al, 2015]
 - RNN + CNN based model



A person on a beach flying a kite.



A black and white photo of a train on a train track.



A person skiing down a snow covered slope.



A group of giraffe standing next to each other.



RNN Applications: Demo

— — —

- [Generating Polyphonic Music](#)
- [Sentence Generation from Picture](#)
- [Handwriting Generation Demo](#)
- [Sentiment Analysis](#)

Conclusion

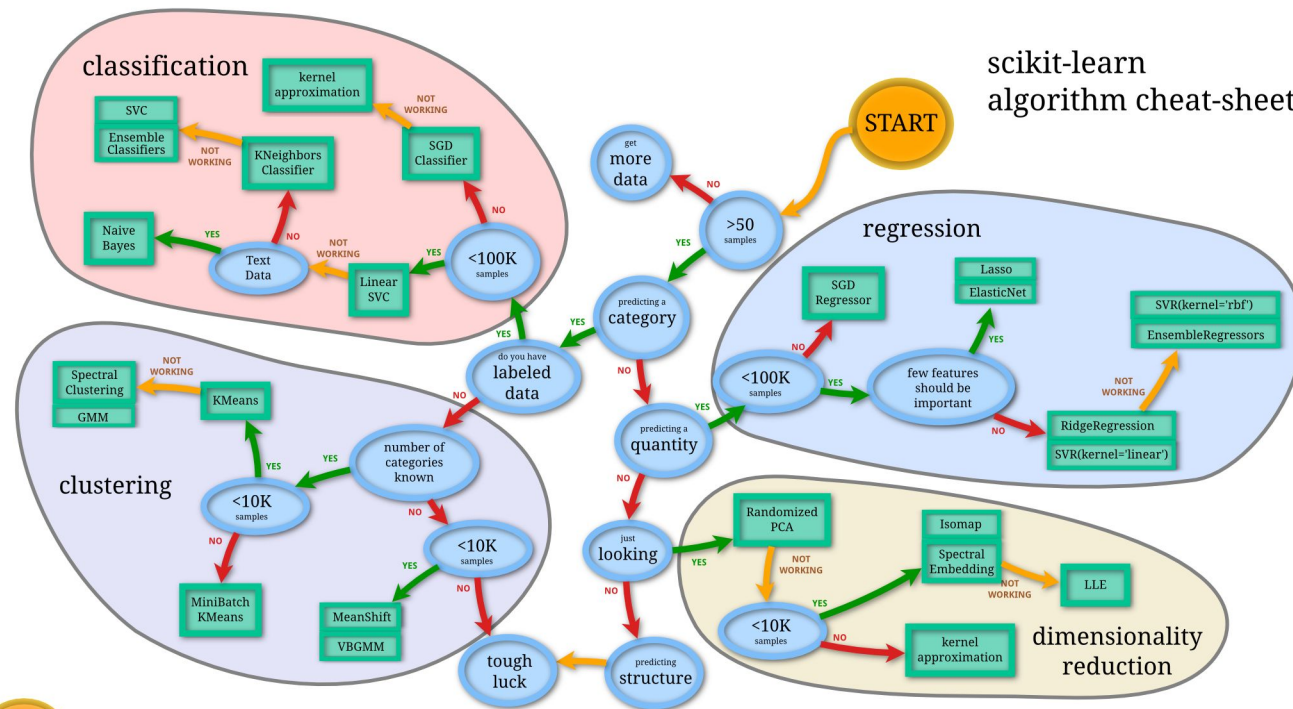
— — —

- Recurrent Neural are able to model sequences
- Training RNN is hard because the vanishing problem
- LSTM tackle the Long-Term Dependency Problem
- Mostly useful in NLP related problems

Useful Resources

ML Cheat-Sheet

scikit-learn
algorithm cheat-sheet

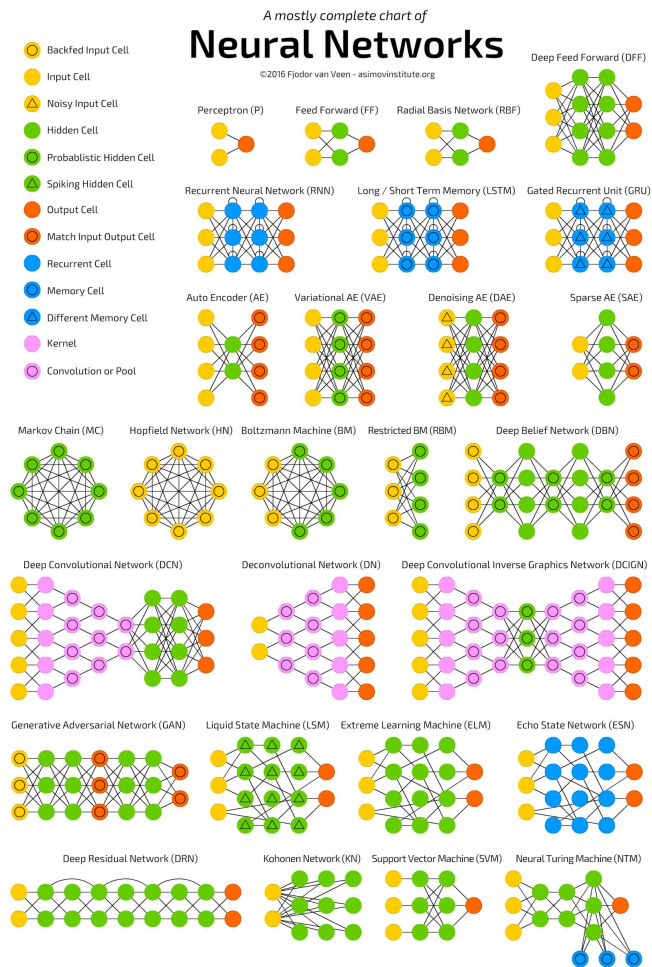


ML Mind Map



The Neural NW Zoo

— — —



Thanks for the Attention

Massimiliano Ruocco, PhD

*Research Scientist, Telenor Research
Adjunct Associate Professor, DART Group, NTNU*



Exercises

Exercise

https://github.com/ruoccoma/ml_talk_hioa