SAPIENZA
UNIVERSITÀ DI ROMA

Elements of Seismology & Machine Learning

Daniele Trappolini
daniele.trappolini@uniroma1.it

Giacomo Mastella
giacomo.mastella@uniroma1.it

Opening

Table of contents

01.

Introduction

This section serves as an introduction to our comprehensive exploration of artificial intelligence. We'll provide an overview of key concepts, history, and the impact of AI and machine learning technologies in various sectors.

A.I. TIMELINE

1950

TURING TEST

Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking it is human, then it has intelligence

1955

A.I. BORN

Term 'artificial intelligence' is coined by computer scientist, John McCarthy to describe "the science and engineering of making intelligent machines"

1961

UNIMATE

First industrial robot, Unimate, goes to work at GM replacing humans on the assembly line

1964

ELIZA

Pioneering chatbot developed by Joseph Weizenbaum at MIT holds conversations with humans

1966

SHAKEY

The 'first electronic person' from Stanford, Shakey is a general-purpose mobile robot that reasons about its own actions

A.I. WINTER

Many false starts and dead-ends leave A.I. out in the cold

1997

DEEP BLUE

Deep Blue, a chess-playing computer from IBM defeats world chess champion Garry Kasparov

1998

KISMET

Cynthia Breazeal at MIT introduces KISmet, an emotionally intelligent robot insofar as it detects and responds to people's feelings

AlphaGo

1999

AIBO

Sony launches first consumer robot pet dog AiBO (AI robot) with skills and personality that develop over time

2002

ROOMBA

First mass produced autonomous robotic vacuum cleaner from iRobot learns to navigate and clean homes

2011

SIRI

Apple integrates Siri, an intelligent virtual assistant with a voice interface, into the iPhone 4S

2011

WATSON

IBM's question answering computer Watson wins first place on popular $1M prize television quiz show *Jeopardy*

2014

EUGENE

Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is human

2014

ALEXA

Amazon launches Alexa, an intelligent virtual assistant with a voice interface that completes shopping tasks

2016

TAY

Microsoft's chatbot Tay goes rogue on social media making inflammatory and offensive racist comments

2017

ALPHAGO

Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number (2^{170}) of possible positions

AI Research

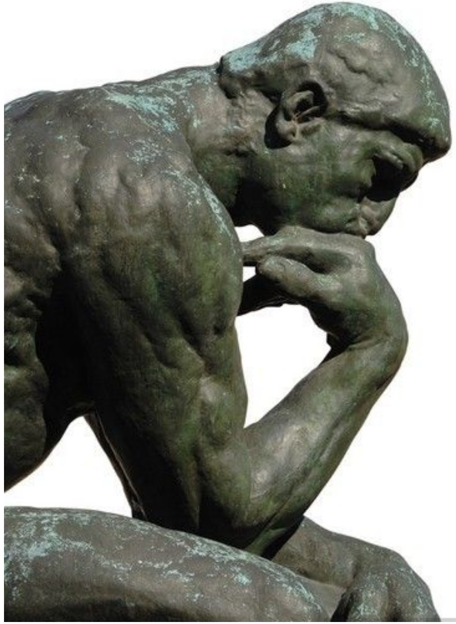
Nobel Prize in Physics 2024
John Hopfield and Geoffrey Hinton

Nobel Prize in Chemistry 2024
David Baker, Demis Hassabis e John M. Jumper

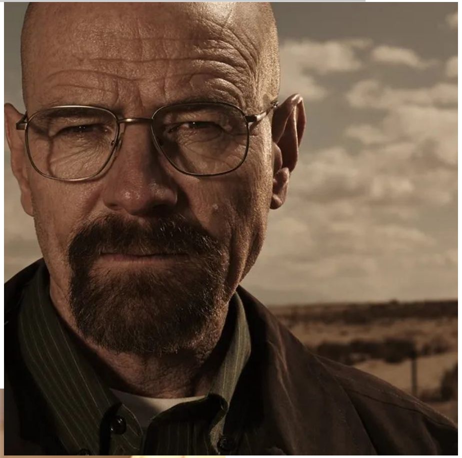
"None of them are chemists or physicists; they are all data scientists."

Introduction

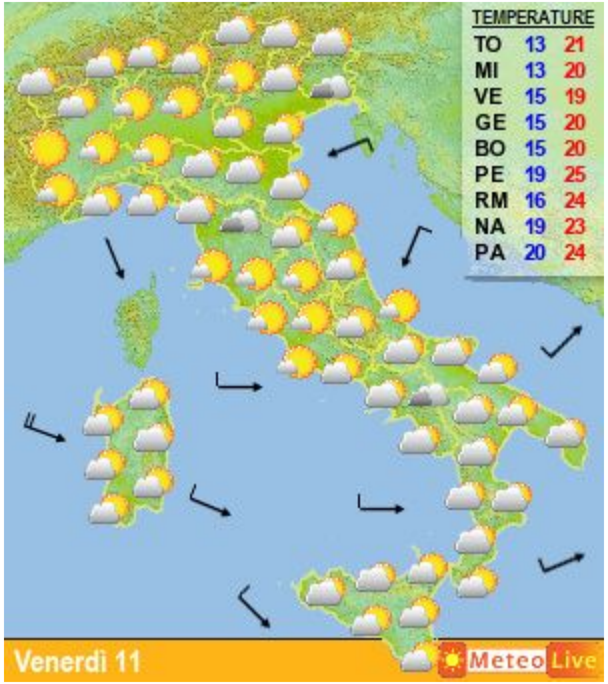
AI Research

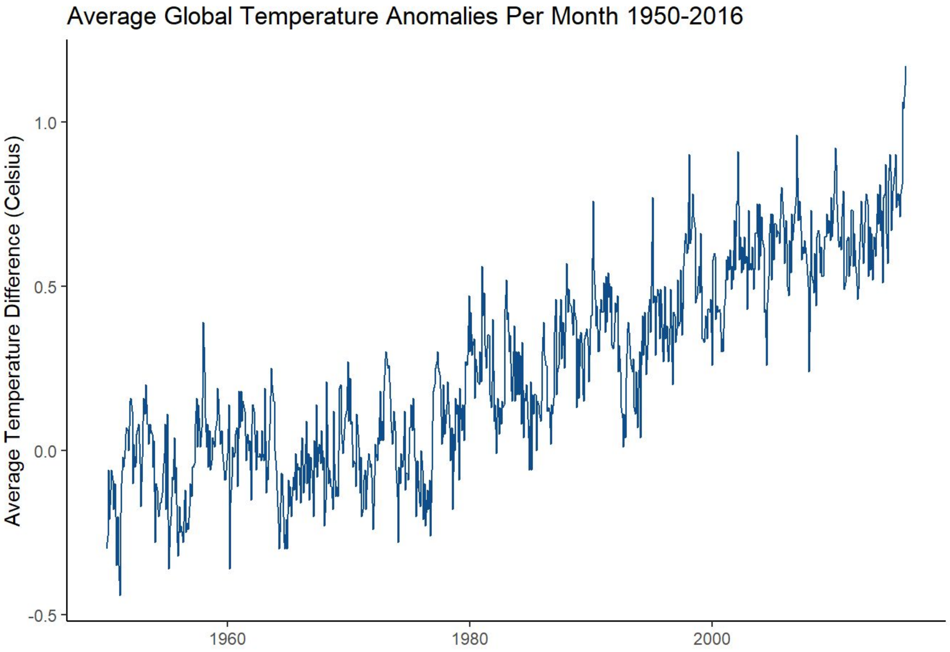


How they do that?



How do we see the world?

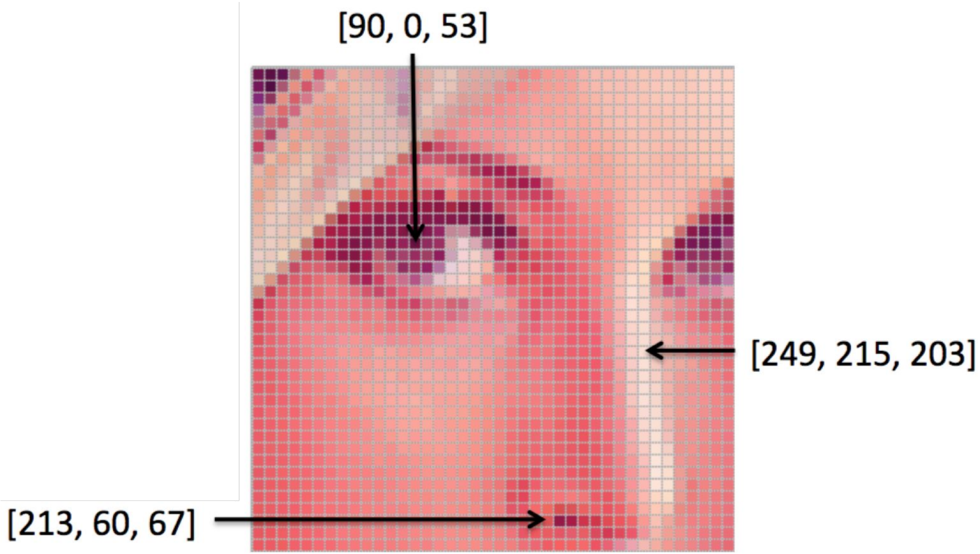




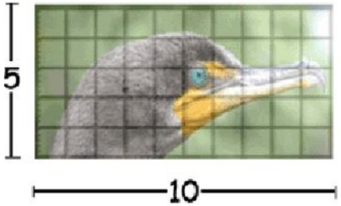
VECTOR

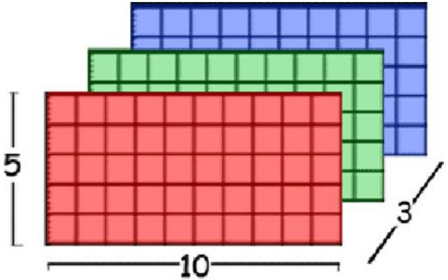
Introduction

How do we see the world?



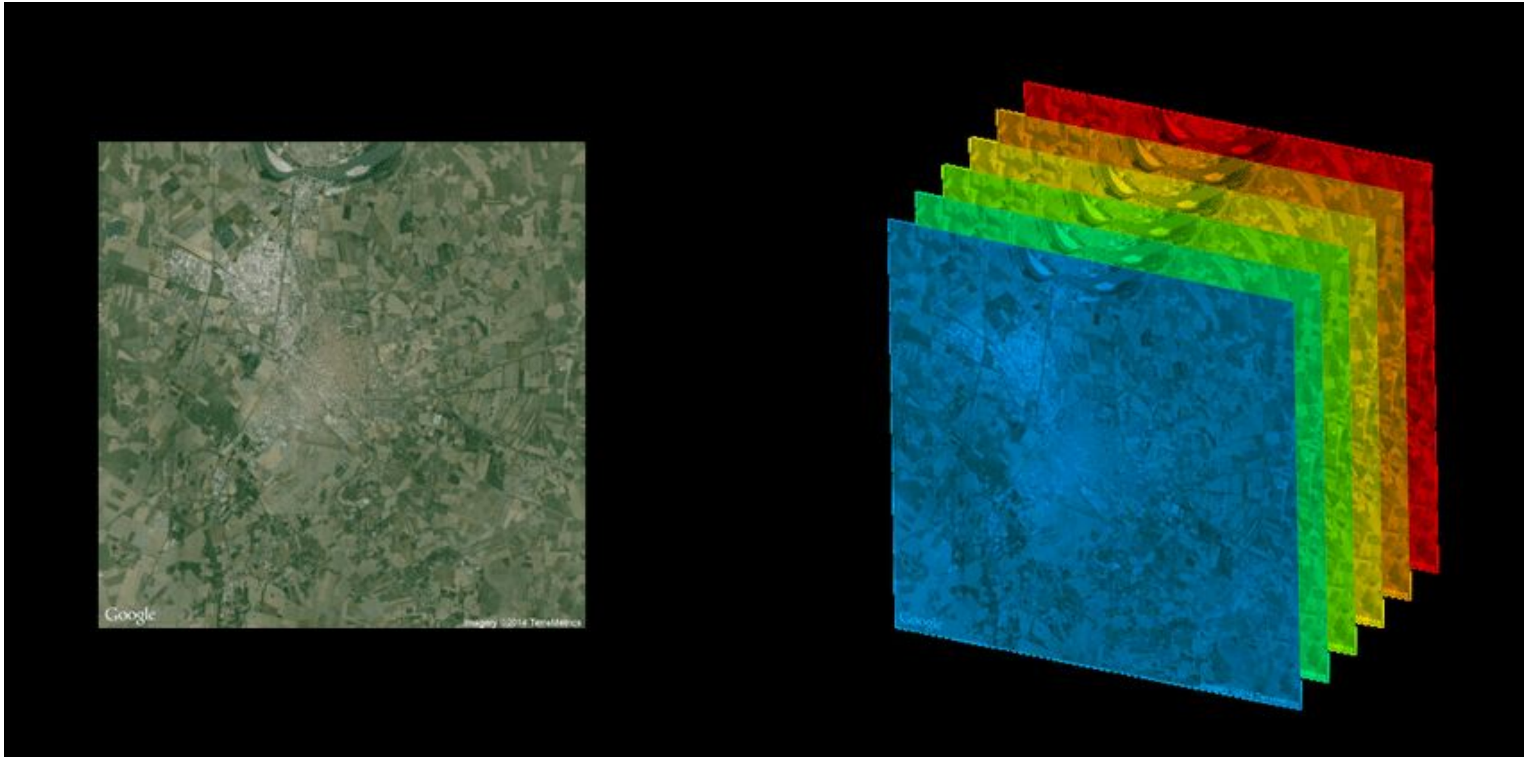
MATRIX



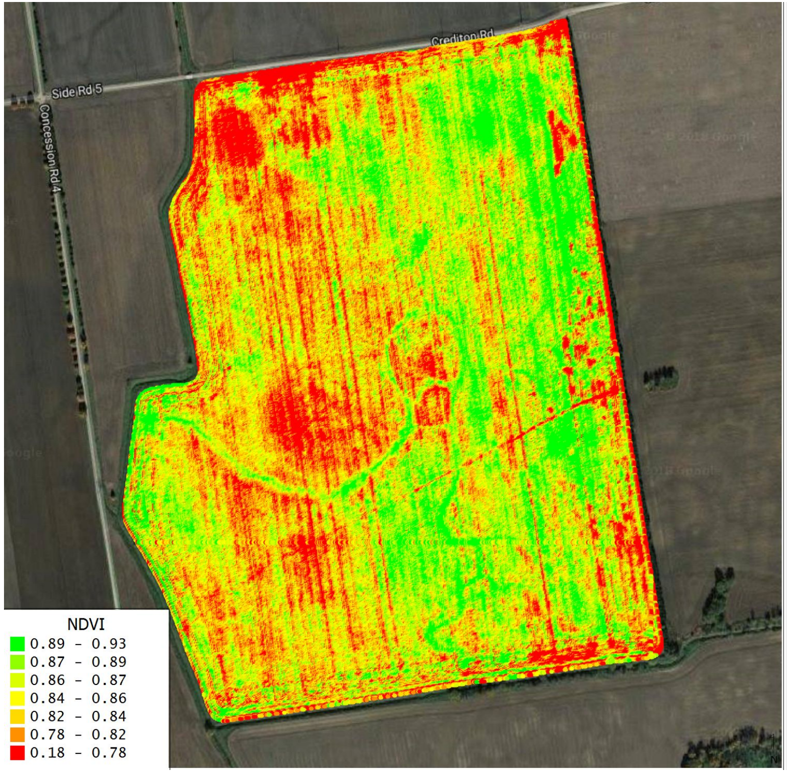


Introduction

How do we see the world?



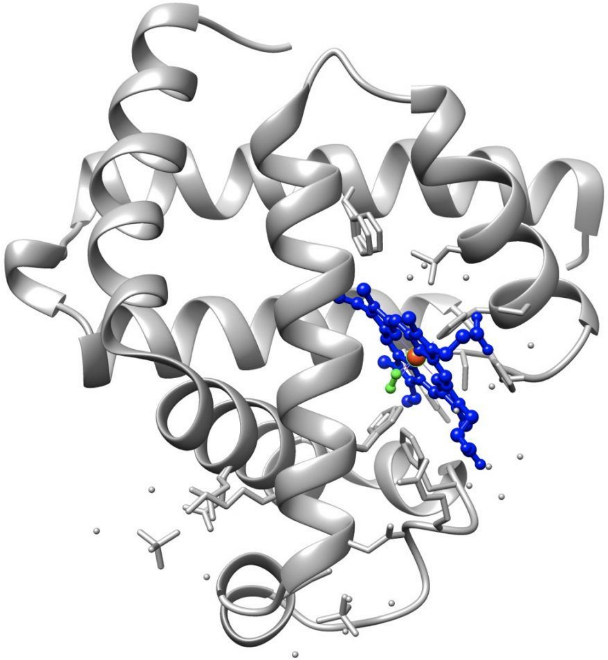
How do we see the world?

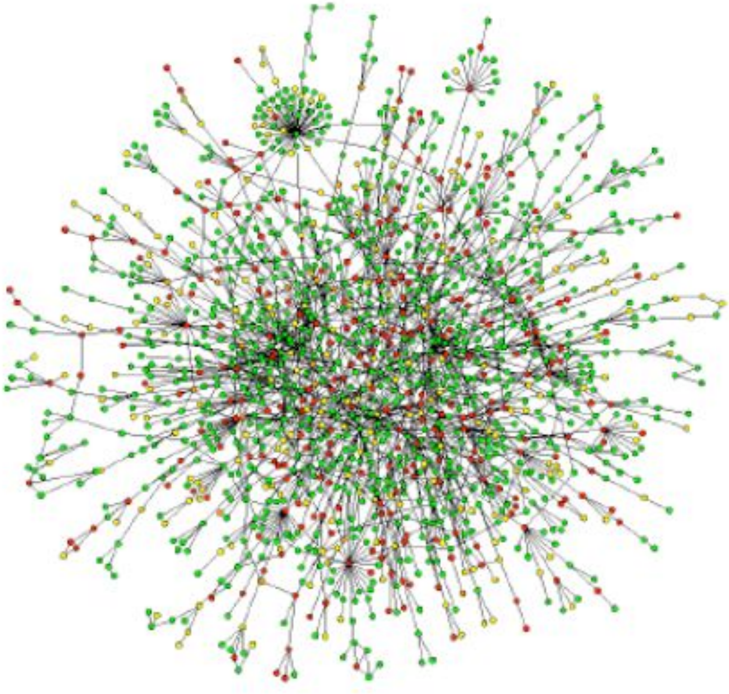


How do we see the world?

TENSOR

How do we see the world?

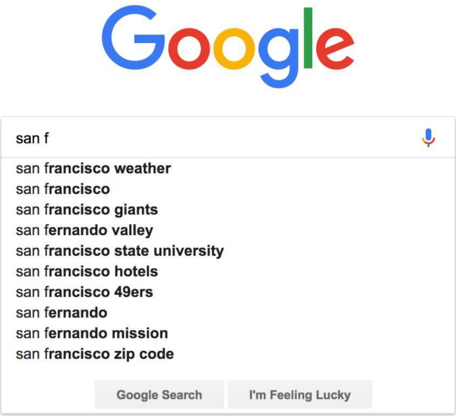




GRAPH

Current AI model development

Text to text: ***ChatGPT***



December 10, 2004

November 30, 2022

"Tell me about San Francisco"

Text to image: ***Midjourney***

October 15, 2022

March 05, 2023

"A young woman with vibrant red hair and striking blue eyes stands amidst a gentle snowfall, medieval-inspired armor, …"

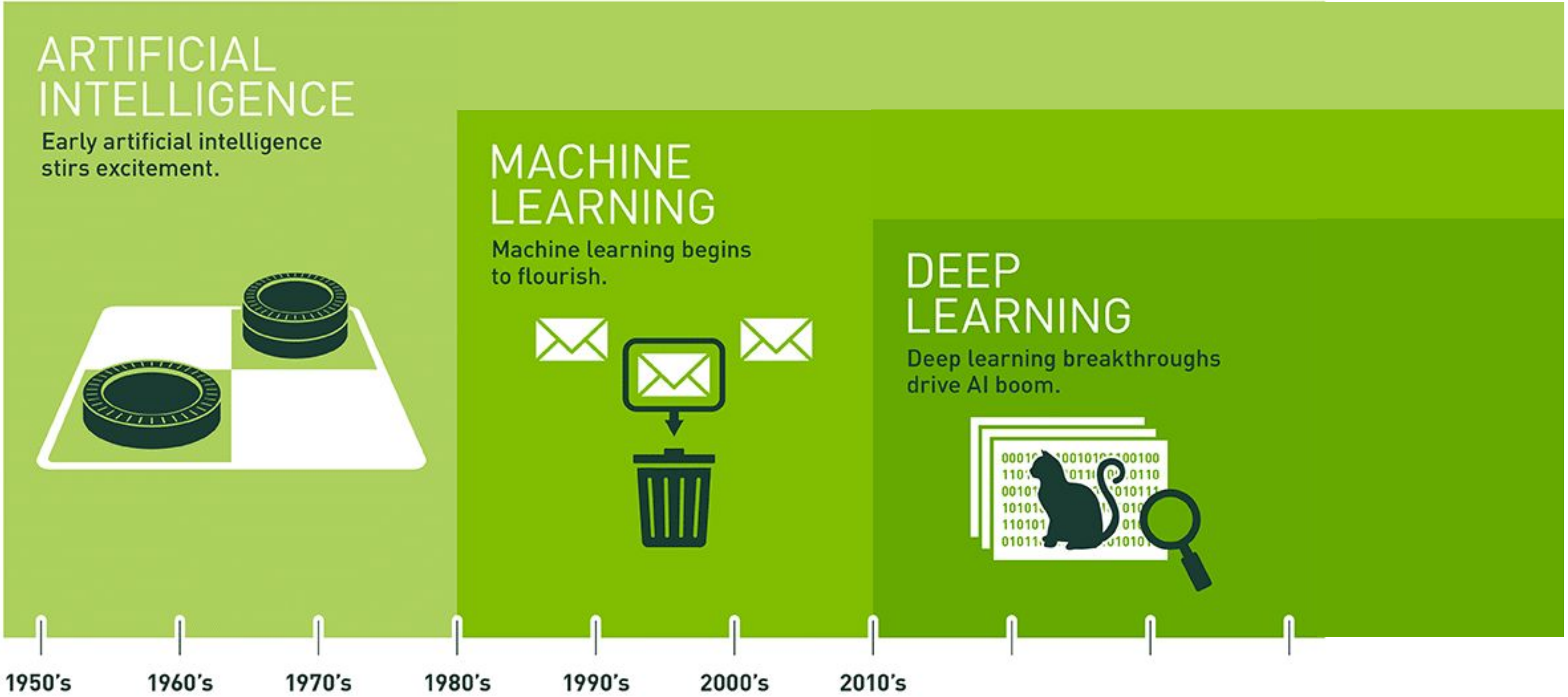
Text to video: ***Sora***

March 30, 2023

February 15, 2024

"A movie trailer featuring the adventures of the 30 year old space man wearing a red wool knitted motorcycle helmet, blue sky, salt desert, cinematic style, shot on 35mm film, …"

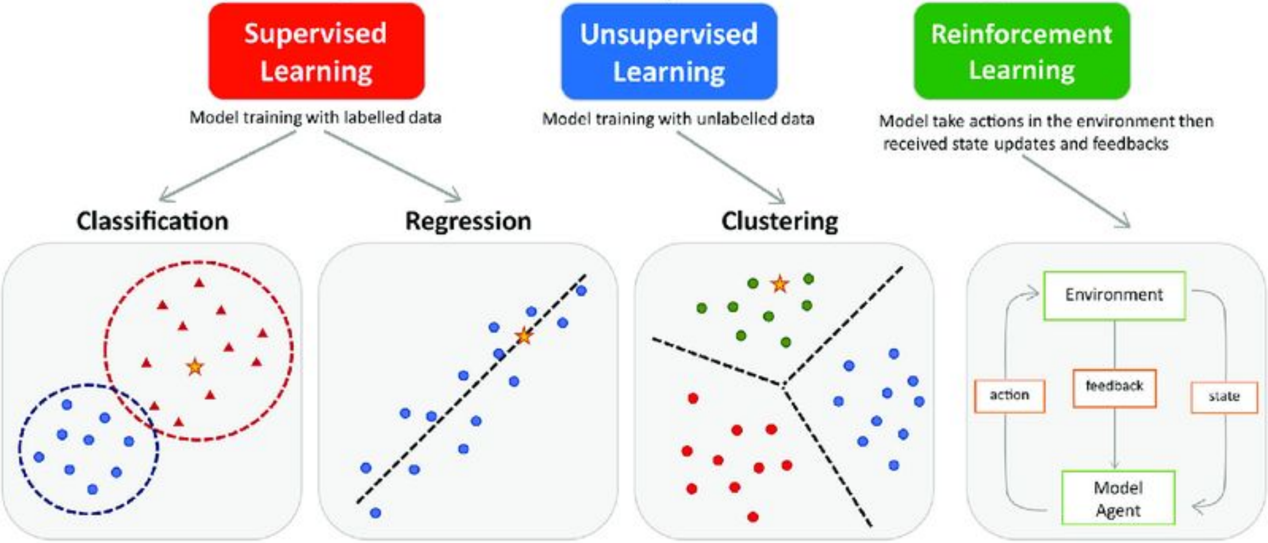
Artificial intelligence



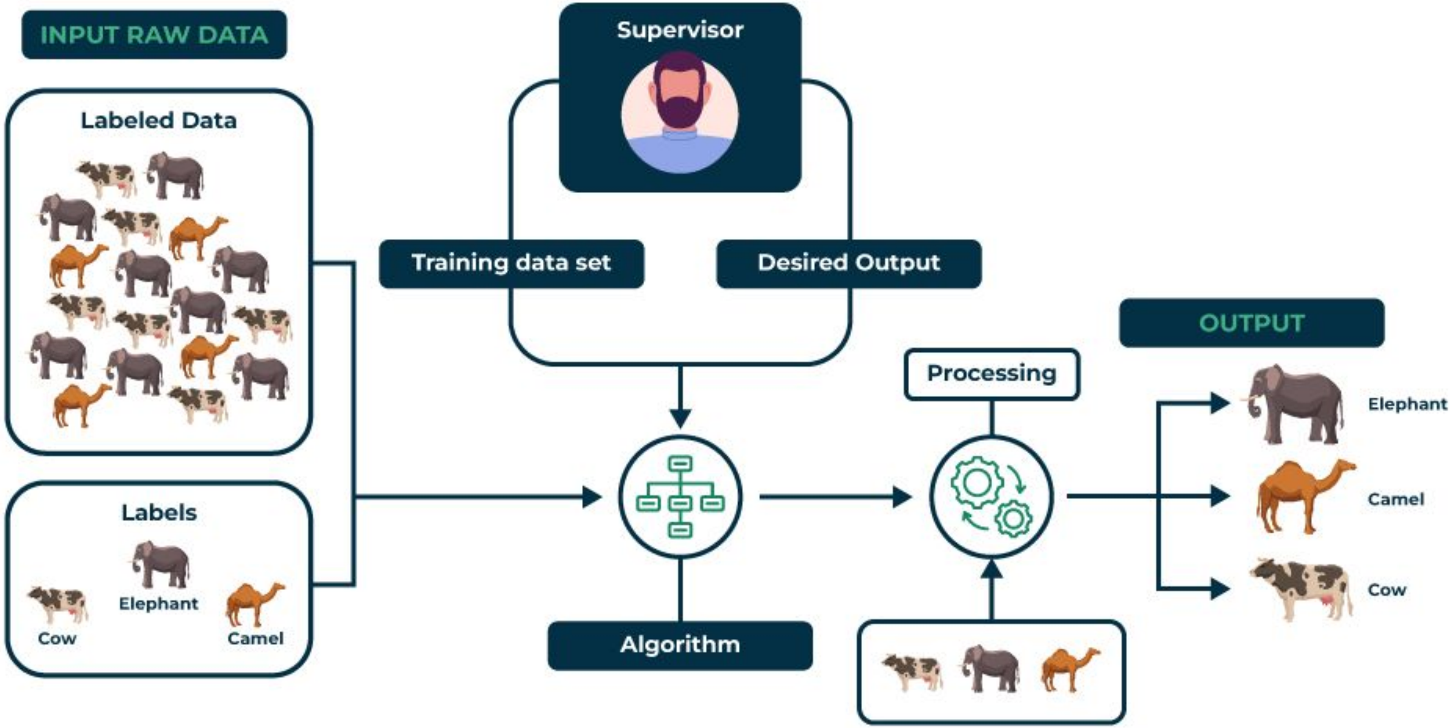
Machine Learning Paradigm

Three main paradigms

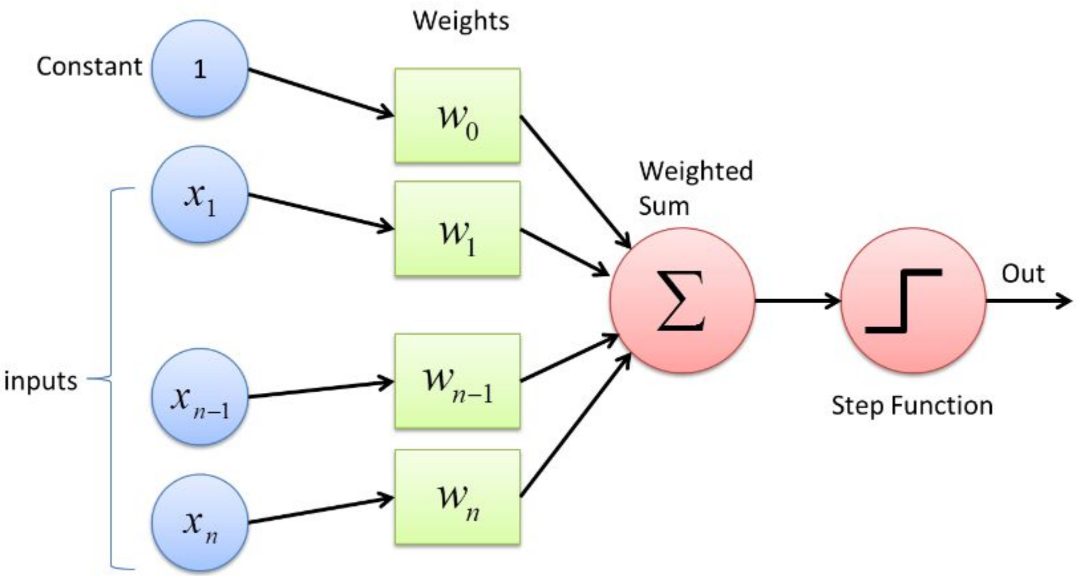
"Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed" ~Arthur Samuel (1959)



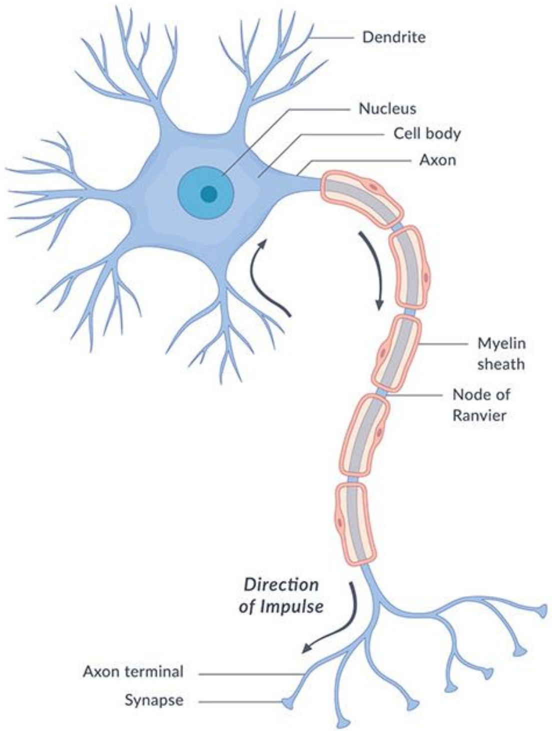
Supervised learning



Supervised learning: ***Perceptron***

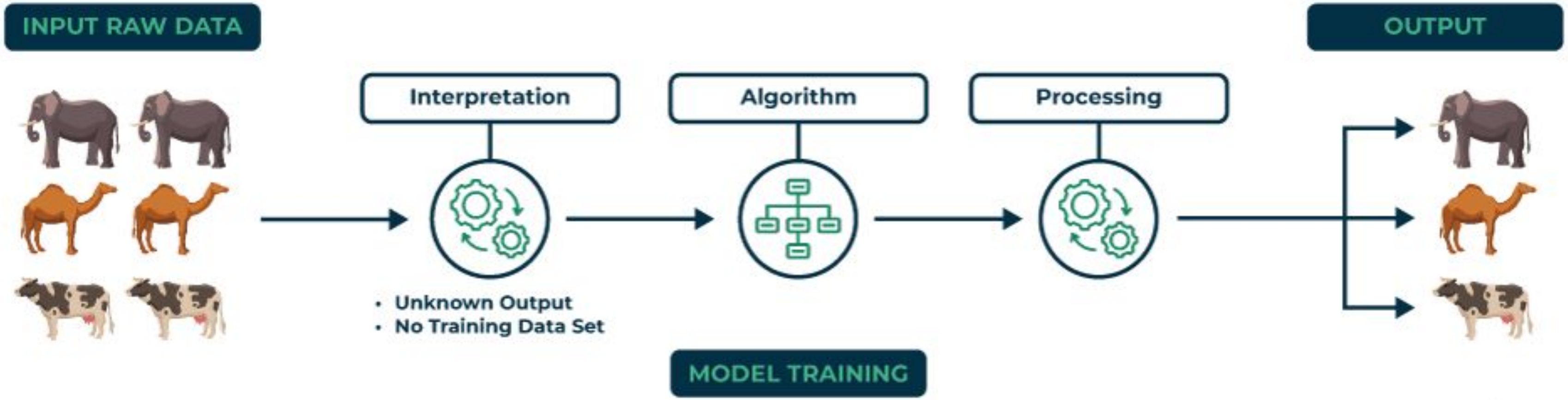


Rosenblatt et al. 1957

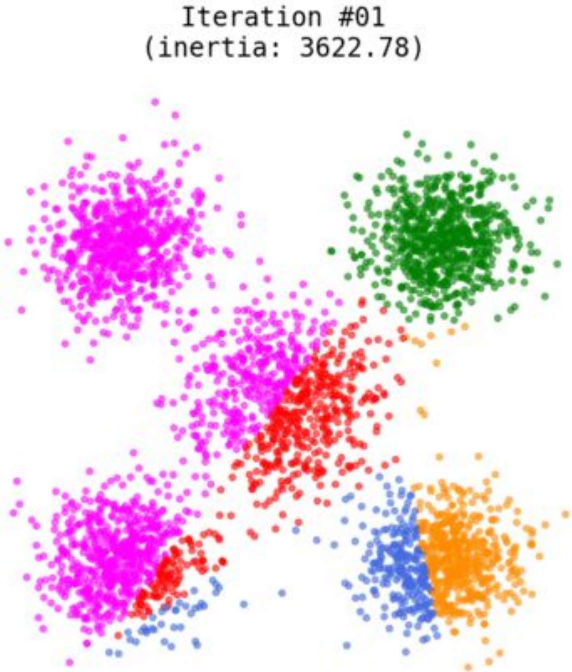


Unsupervised learning

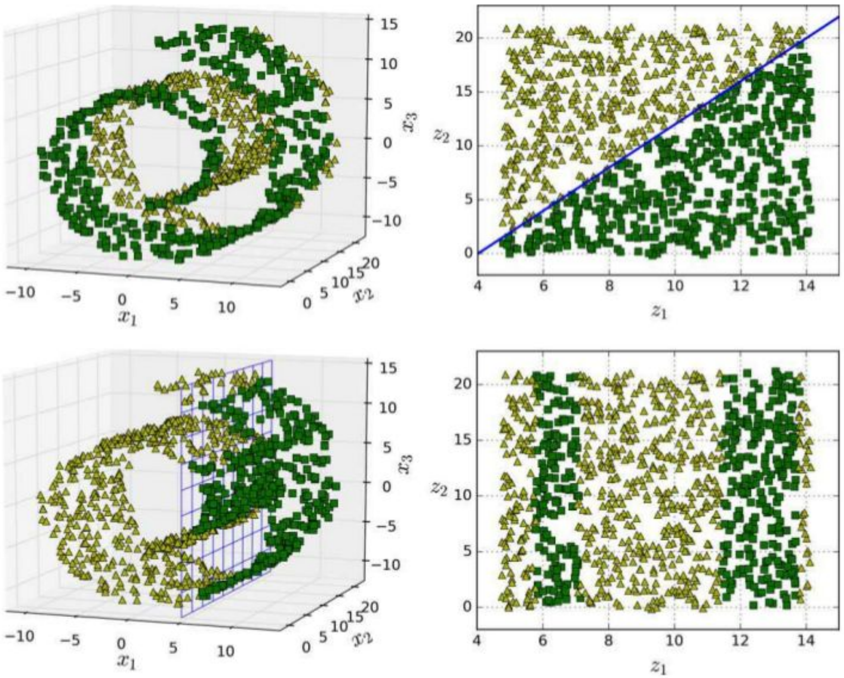
Unsupervised learning is a type of machine learning where models discover patterns in data without pre-existing labels, often used for clustering or dimensionality reduction.



Unsupervised learning: ***K-Means / PCA***



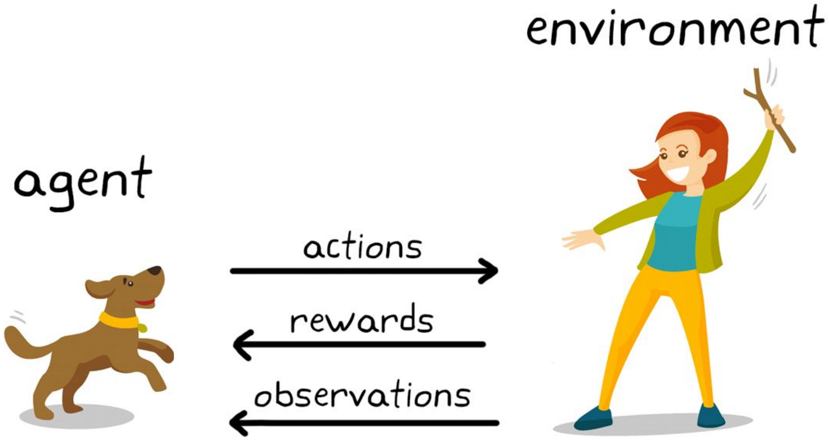
K-Means



PCA

Reinforcement learning

Reinforcement Learning is a machine learning method where an agent learns optimal actions through trial and error to maximize rewards in an environment.



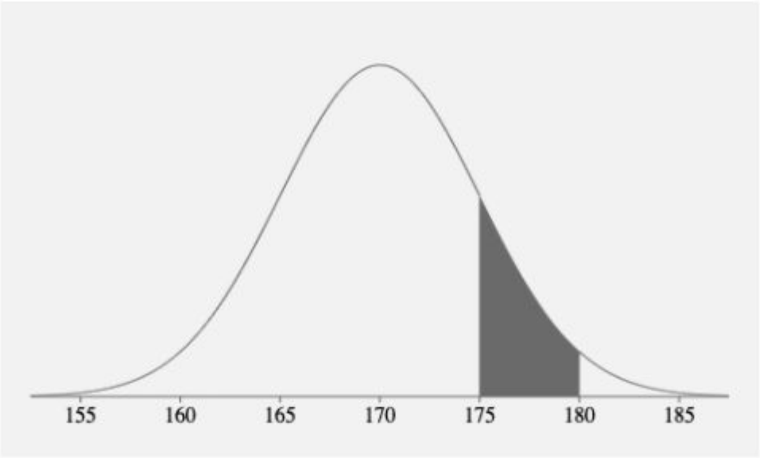
02.

Linear Regression

In this section, we focus on **Linear Regression**, one of the fundamental techniques in machine learning used for predictive modeling. Linear regression aims to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. We'll explore how linear regression helps to predict outcomes, analyze trends, and estimate values by learning from the underlying patterns in the data.

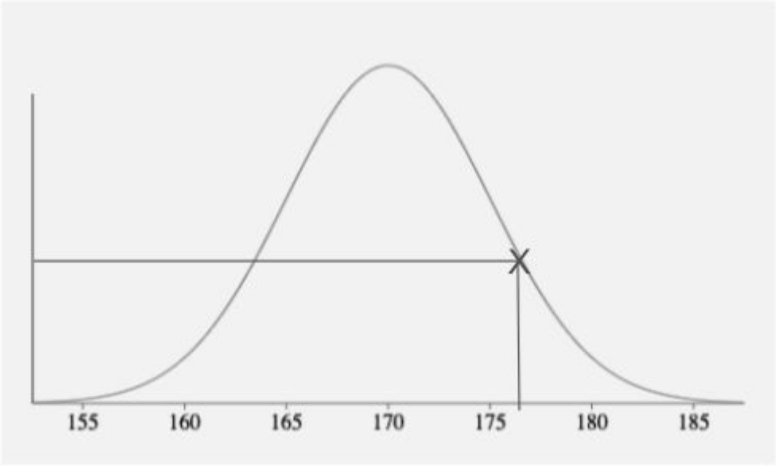
Probability vs Likelihood

Probability



Pr(Data | Distribution)

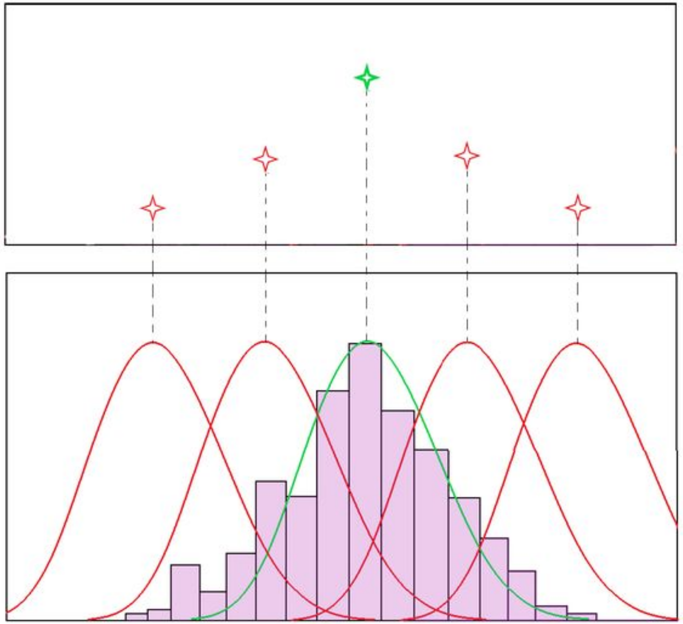
Likelihood



L(Distribution | Data)

Maximum Likelihood Estimator

Maximum likelihood estimate plot



Multiple PDFs over the
random sample histogram plot

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

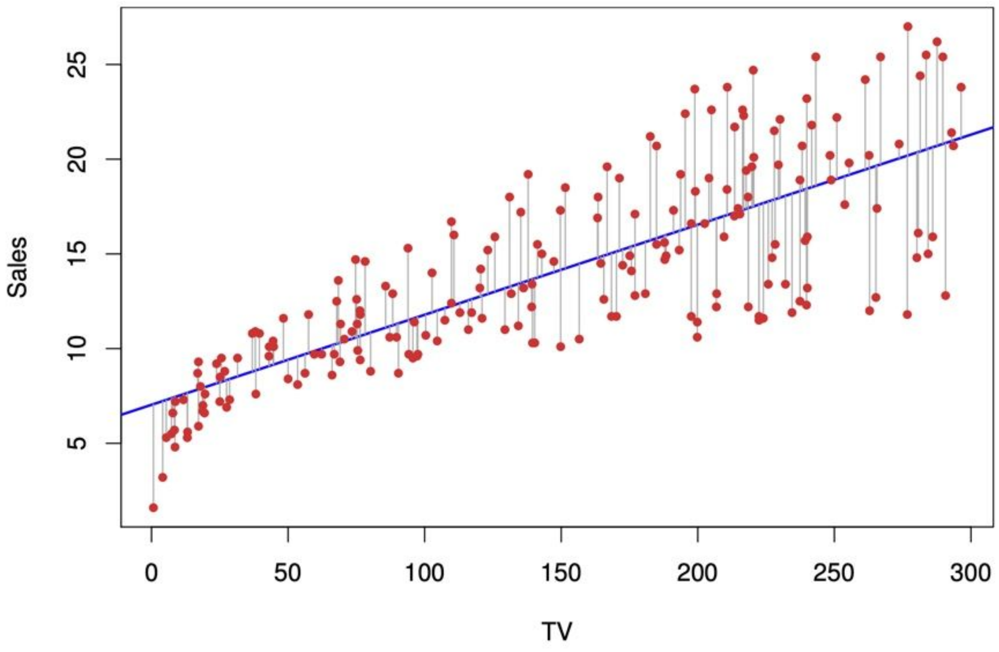
$$\max L(x, \theta) = L(x, \hat{\theta})$$

$$\ell(x, \theta) = \ln \left[\prod_{i=1}^{n} f(x_i, \theta) \right] = \sum_{i=1}^{n} \ln f(x_i, \theta)$$

$$\frac{\partial \ell(x, \theta)}{\partial \theta} = [L(x, \theta)]^{-1} \frac{\partial L(x, \theta)}{\partial \theta}$$

Linear Regression

Ordinary Least Square



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}))^2$$

TEOREMA GAUSS-MARKOV

$$\mathbb{E}[\varepsilon] = 0,$$

$$\mathrm{Var}(\varepsilon) = \sigma^2 I,$$

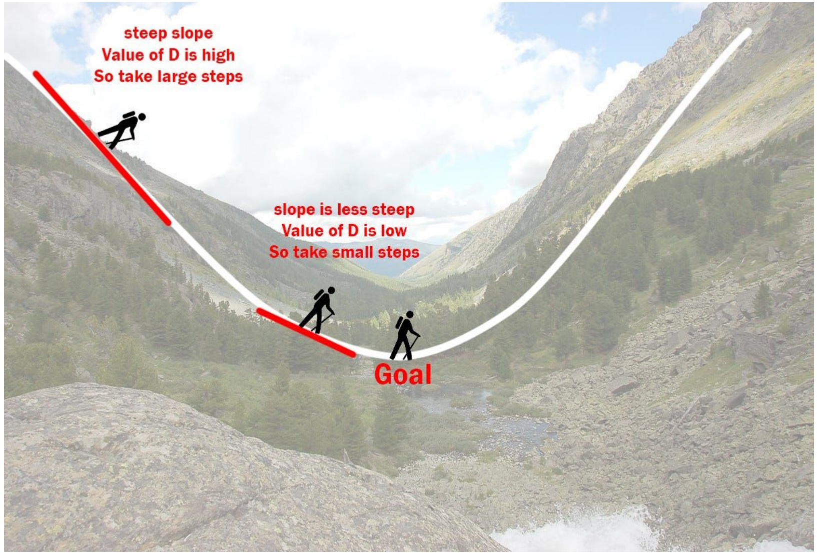
$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

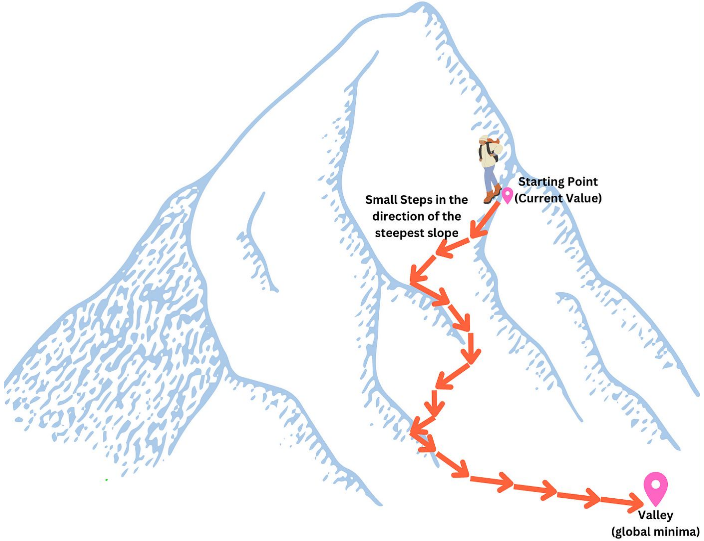
$\Rightarrow$ **BLUE**

Gradient Descent

Gradient descent

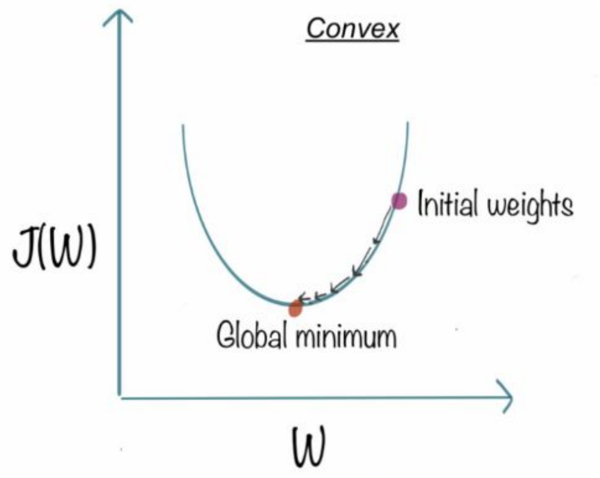
Is an optimization algorithm that iteratively adjusts parameters to minimize a cost function, moving in the direction of steepest decrease.

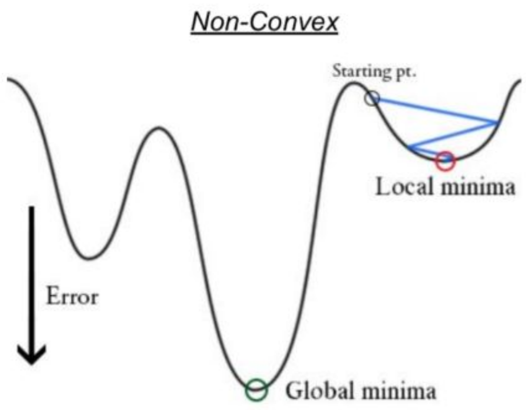




Cost (or loss) function

A mathematical function that measures the difference between the algorithm's predictions and the actual data. It guides the optimization process by quantifying the model's performance.





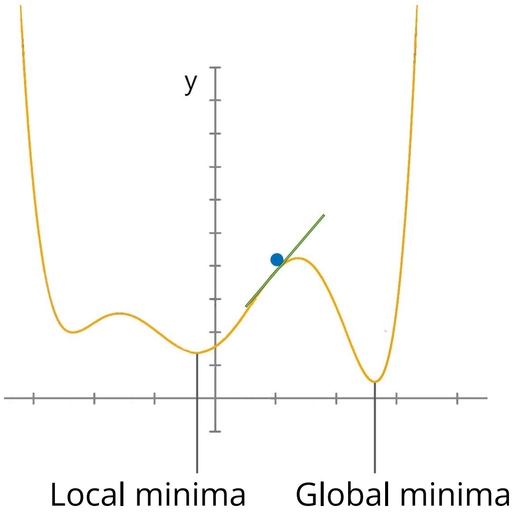
$$\theta := \theta - \alpha \nabla J(\theta)$$

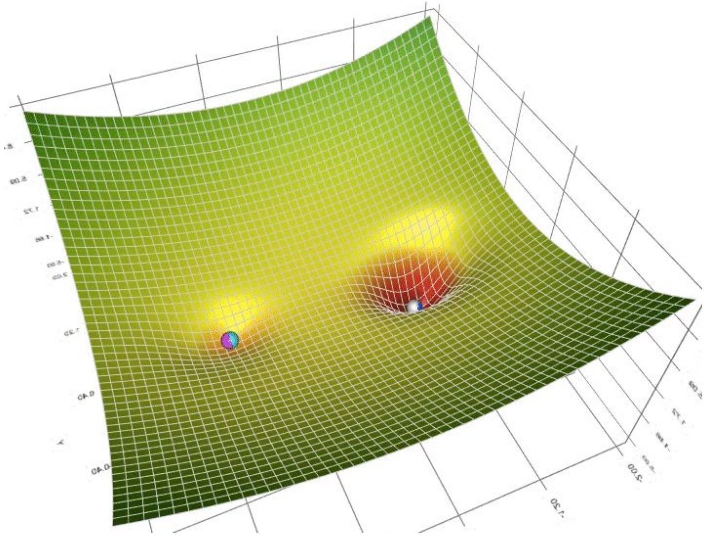
$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

Training

Gradient descent

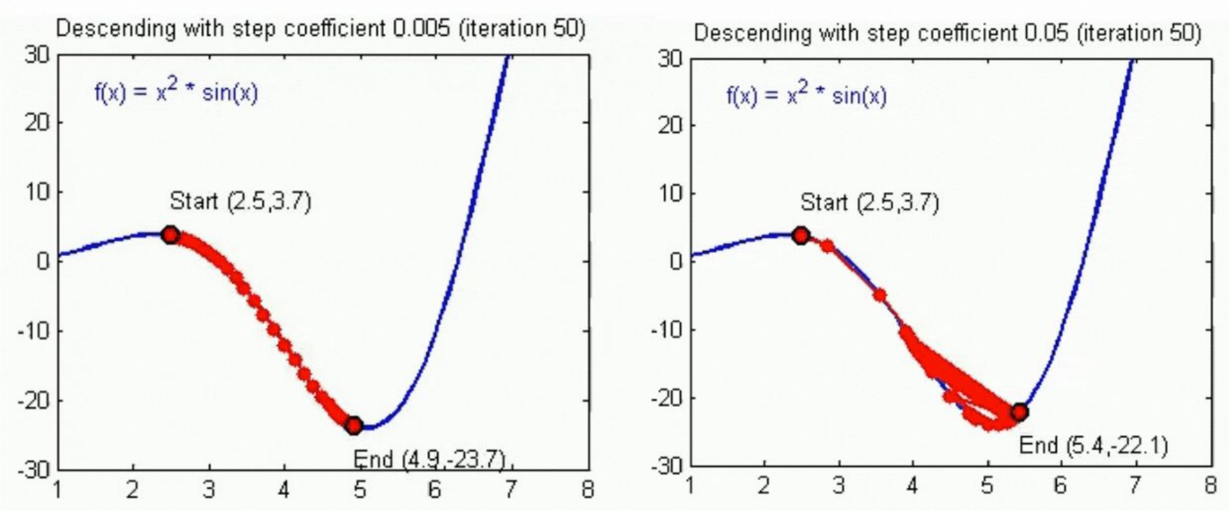
Is an optimization algorithm that iteratively adjusts parameters to minimize a cost function, moving in the direction of steepest decrease.





Learning rate

An hyperparameter that controls the adjustment of model weights during training. It determines the size of the steps the algorithm takes to reach the minimum of the loss function.



Evaluate Regression Model

Evaluate Regression

```
Call:
lm(formula = height ~ age, data = ageandheight)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27238 -0.24248 -0.02762  0.16014  0.47238 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  64.9283     0.5084  127.71  < 2e-16 ***
age           0.6350     0.0214   29.66 4.43e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.256 on 10 degrees of freedom
Multiple R-squared:  0.9888,    Adjusted R-squared:  0.9876 
F-statistic:   880 on 1 and 10 DF,  p-value: 4.428e-11
```

Evaluate Regression

$$H_0 : \beta_1 = 0 \qquad H_a : \beta_1 \neq 0,$$

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$R^2_{\text{adjusted}} = 1 - \left(\frac{(1 - R^2)(n-1)}{n - p - 1}\right)$$

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_a : \text{ at least one } \beta_j \text{ is non-zero.}$$

$$F = \frac{(\mathrm{TSS} - \mathrm{RSS})/p}{\mathrm{RSS}/(n - p - 1)},$$

- **Assess if the feature has a relationship with the target variable (y).**
- **Evaluate if the entire model has a relationship with the target variable (y).**
- **Evaluate how much better the model performs compared to the dummy model.**

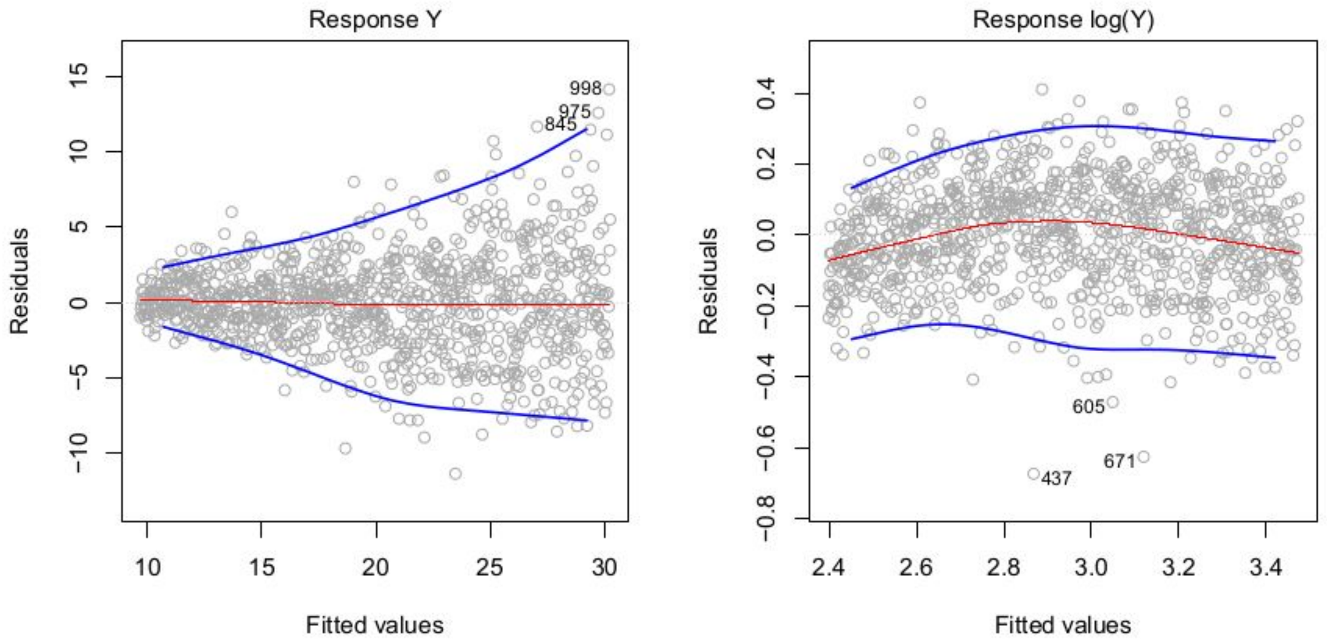
Problem in Regression Model

Linear Regression

GAUSS MARKOV PROBLEM 1

PROBLEM:

ETEROSCHE DASTICITY, non costant residual variance.

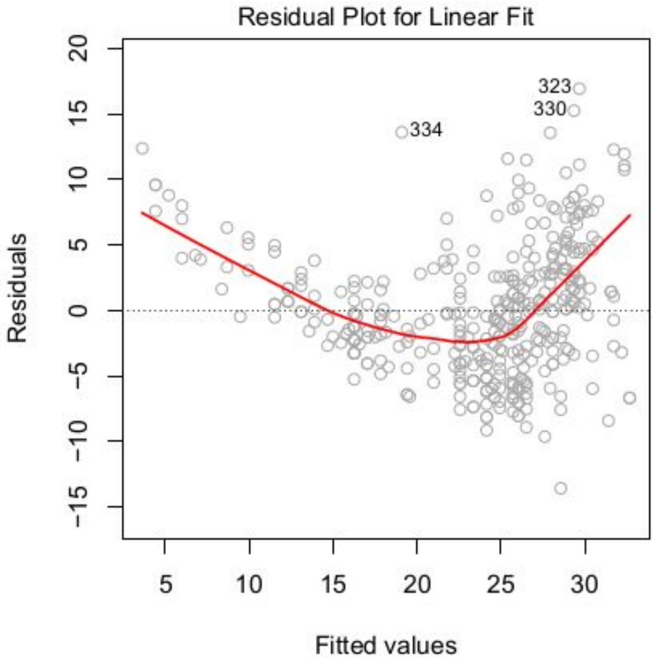


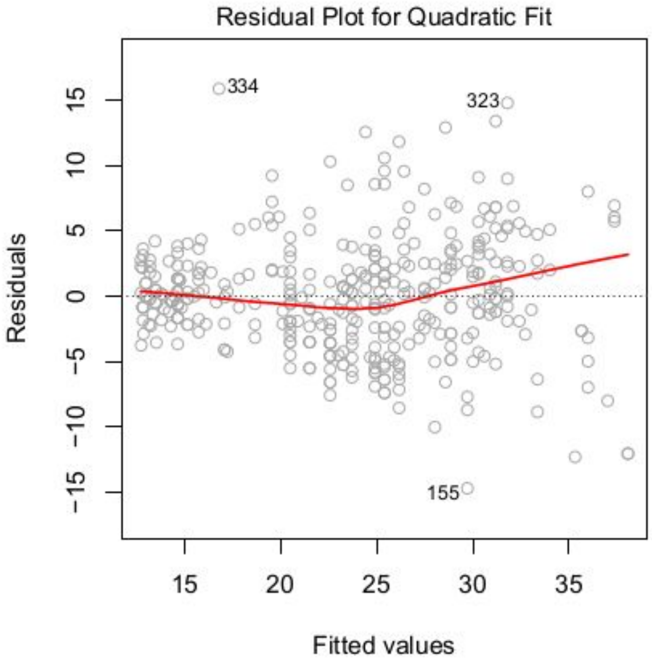
SOLUTION: 1. Y Transformation 2. OLS Weighted Method

GAUSS MARKOV PROBLEM 2

PROBLEM:

NON LINEARITY

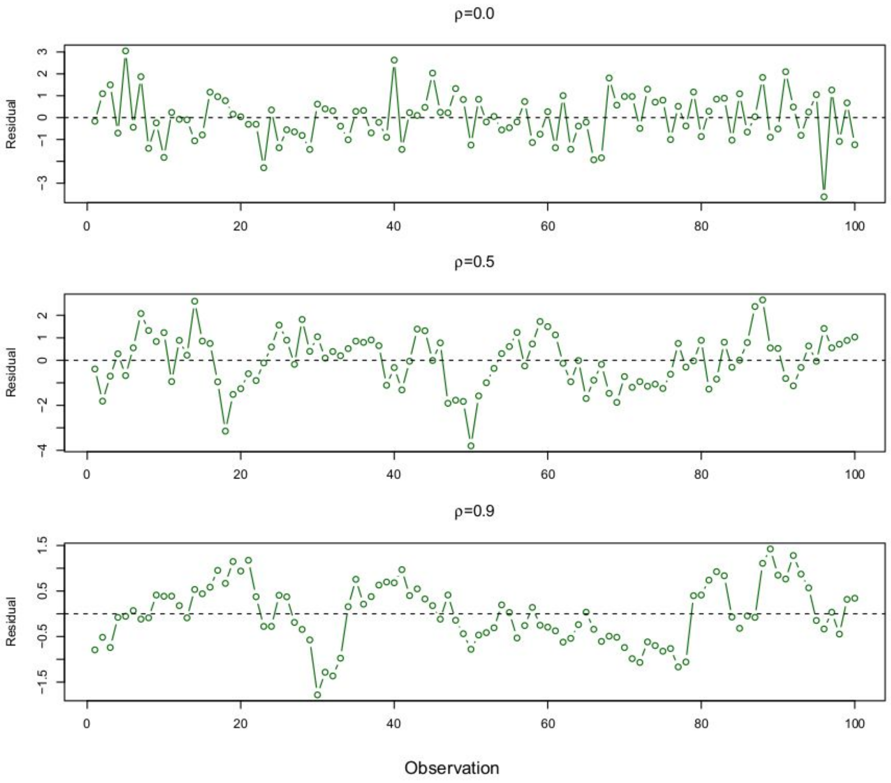




SOLUTION: 1. X Transformation 2. Use an higher polynomial degree

Linear Regression

GAUSS MARKOV PROBLEM 3



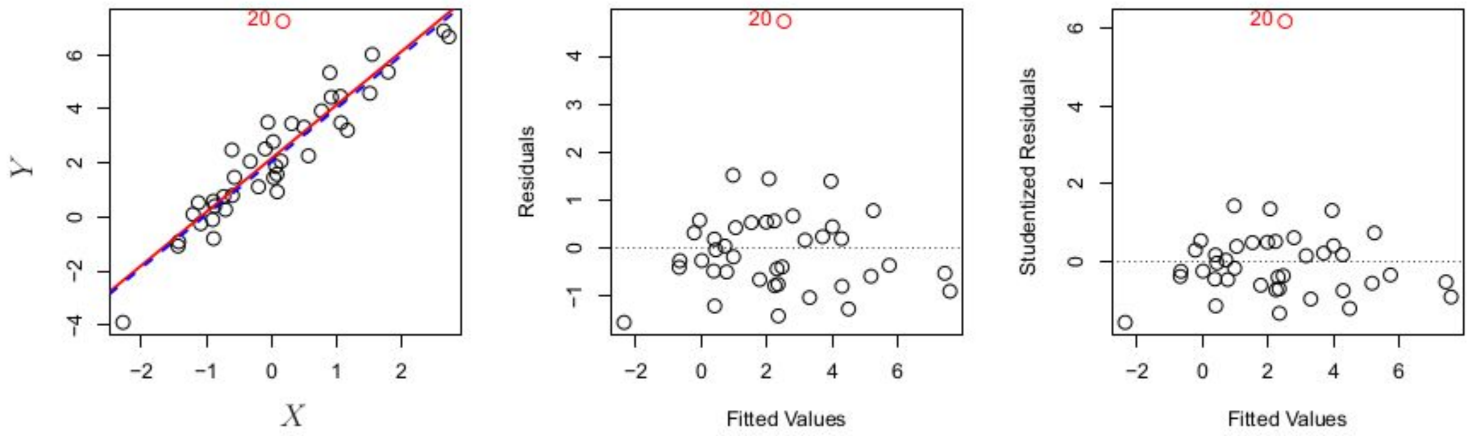
PROBLEM:

COVARIANCE of the Residual not null.

Difficult to detect and to solve, frequently in time series.

SOLUTION:
1. Instrumental Variable

OTHER PROBLEM 4

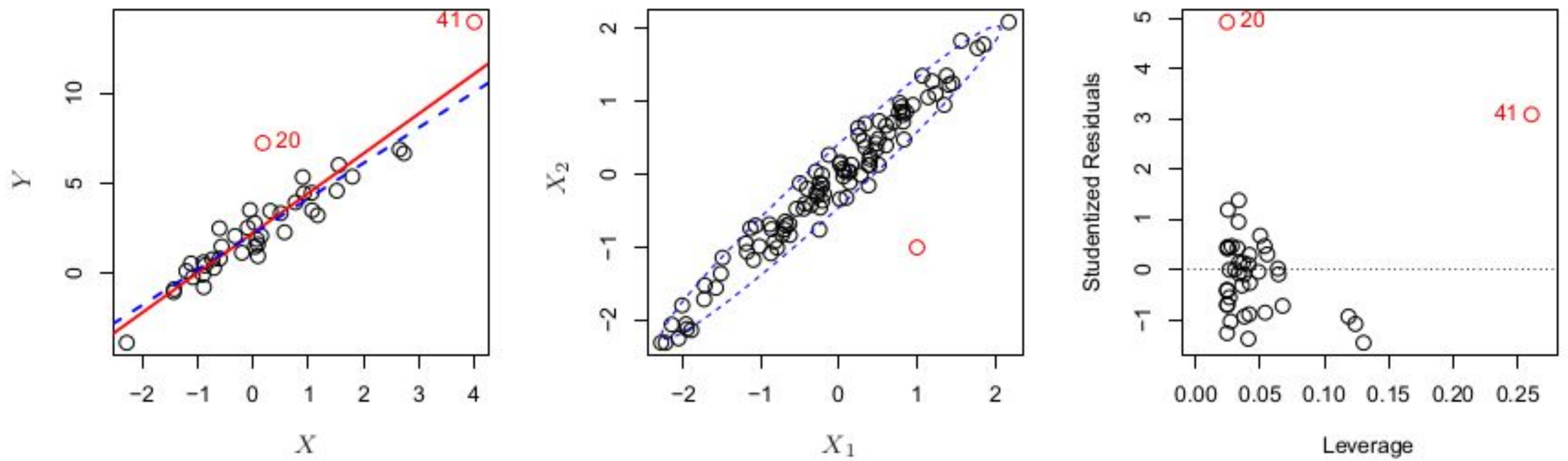


PROBLEM:

OUTLIERS

SOLUTION:
1. Delete Outlier
2. X/Y Trasformation

OTHER PROBLEM 5



PROBLEM:

HIGH LEVERAGE

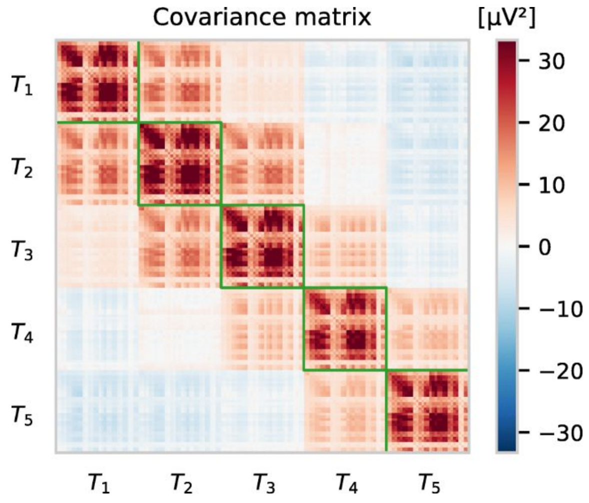
SOLUTION:
1. Delete Outlier

HOW TO DETECT:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i'=1}^{n}(x_{i'} - \bar{x})^2}.$$

Linear Regression

OTHER PROBLEM 6



	Coefficient	Std. error	t-statistic
Intercept	−173.411	43.828	−3.957
age	−2.292	0.672	−3.407
limit	0.173	0.005	34.496
Intercept	−377.537	45.254	−8.343
rating	2.202	0.952	2.312
limit	0.025	0.064	0.384

PROBLEM:

COLLINEARITY AND MULTICOLLINEARITY

HOW TO DETECT:

1. Covariance Matrix
2. $$\mathrm{VIF}(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_{-j}}},$$

SOLUTION:

1. Delete Features

03.

Applications

After a Small BREAK

04.

Logistic Regression

In this section, we focus on Deep Learning, a subset of machine learning that utilizes neural networks with many layers. We'll examine how deep learning models can learn complex patterns and perform tasks like image and speech recognition.

05.

Applications

In this section, we focus on Deep Learning, a subset of machine learning that utilizes neural networks with many layers. We'll examine how deep learning models can learn complex patterns and perform tasks like image and speech recognition.

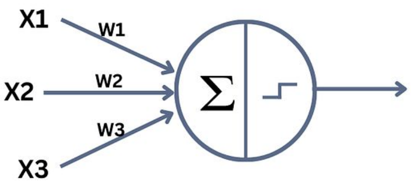
03.

Deep Learning

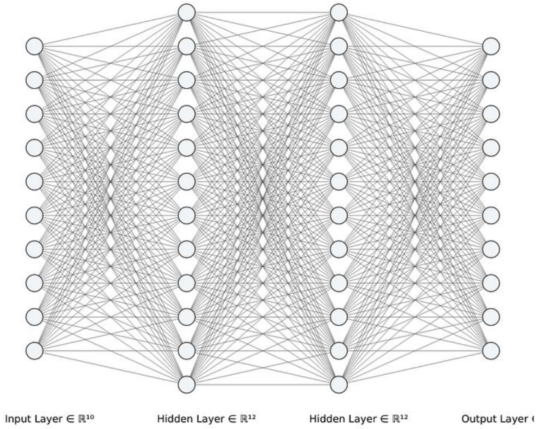
In this section, we focus on Deep Learning, a subset of machine learning that utilizes neural networks with many layers. We'll examine how deep learning models can learn complex patterns and perform tasks like image and speech recognition.

Neural Networks (NNs)

A computational model inspired by the human brain's structure, and consists of layers of interconnected nodes or neurons that process and transmit signals to solve tasks.



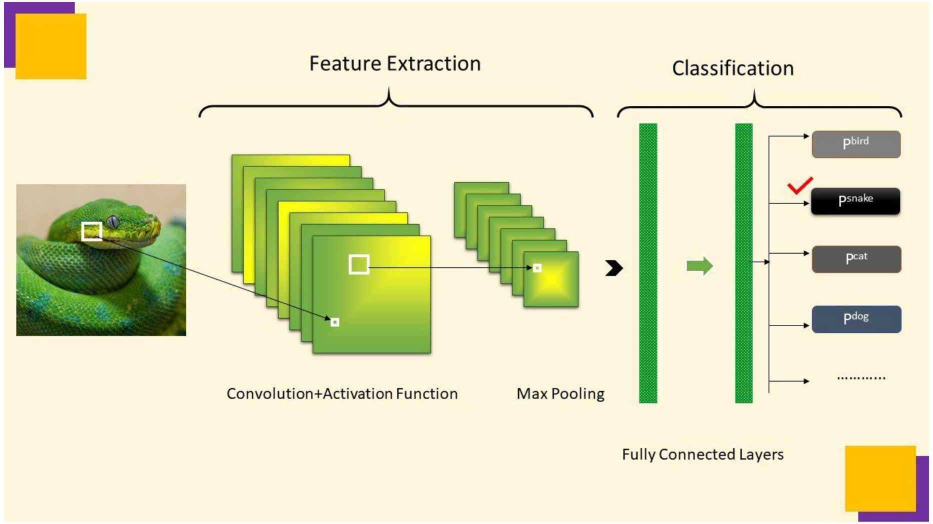
Single-layer perceptron



Multi-layer perceptron

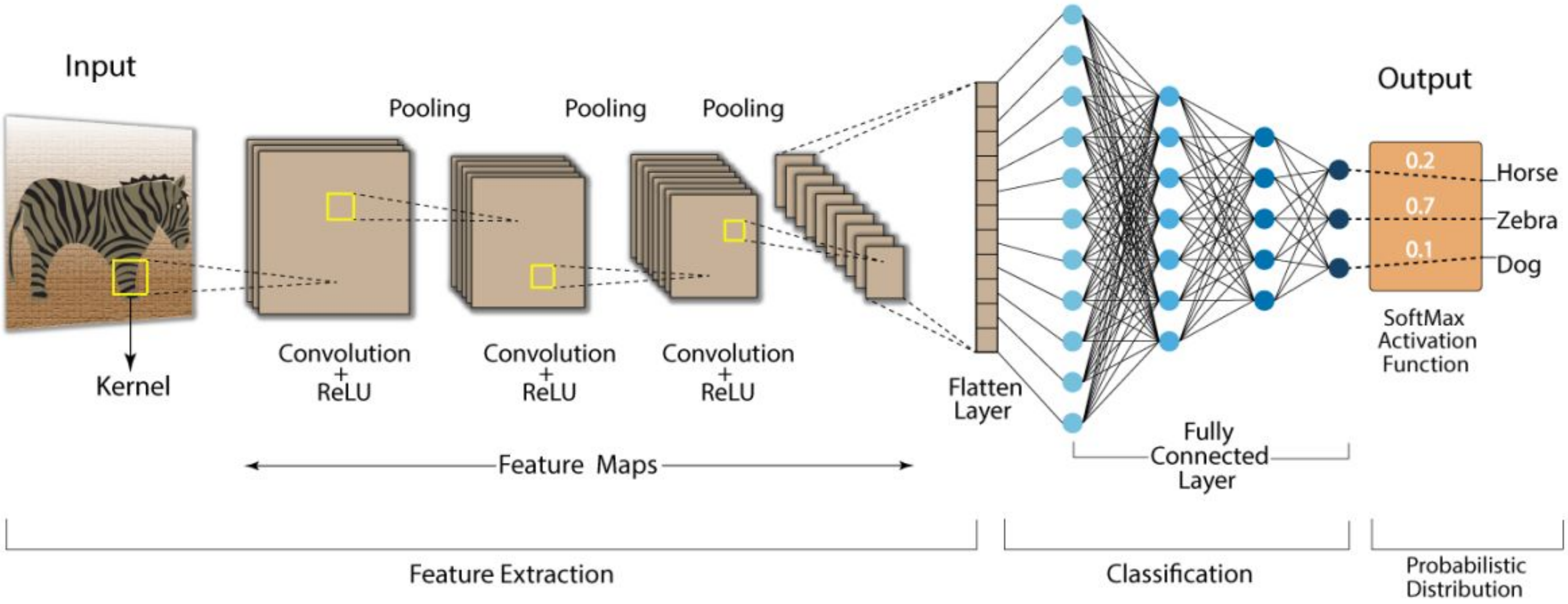
Convolutional Neural Networks: ***What is a convolution***

Convolution involves sliding a smaller array, known as a kernel or filter, over a larger array (the input signal or image) to produce a new array called the convolved feature or feature map.

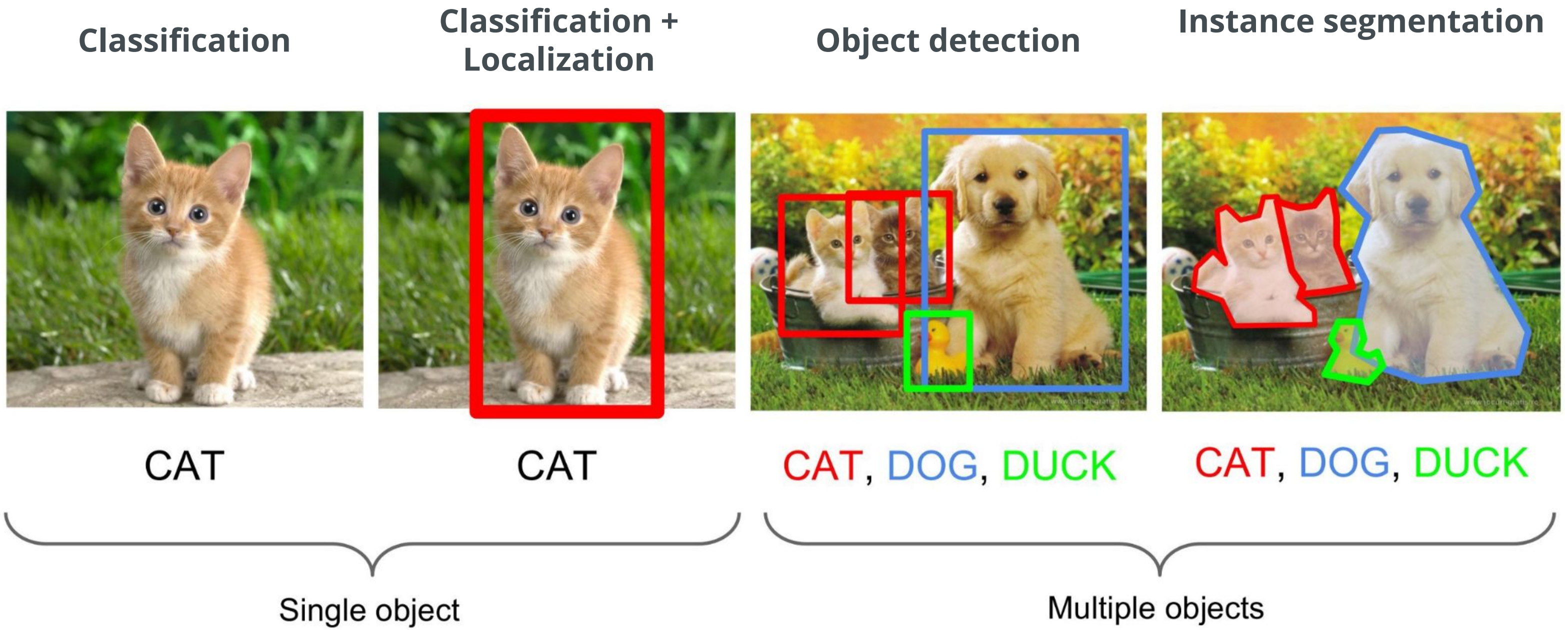


Deep Learning

Convolutional Neural Networks (CNNs)

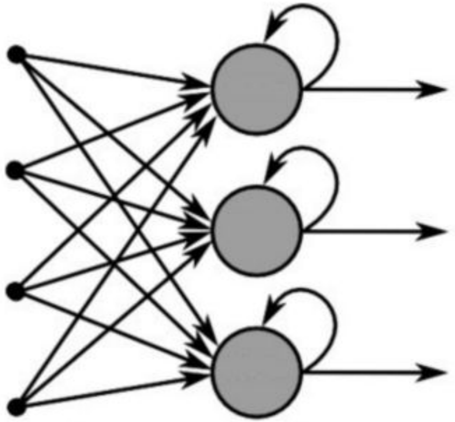


Convolutional Neural Networks: ***Task***

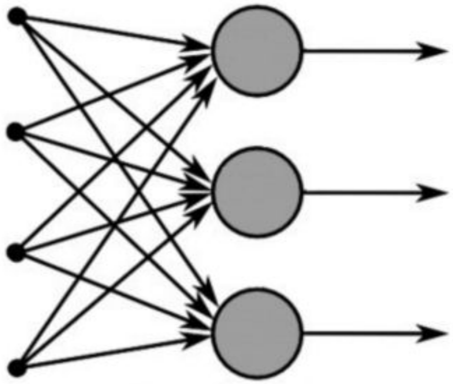


Recurrent Neural Network (RNNs)

CNNs struggle with *time series* because they don't naturally keep track of the order of things. They treat input data as if all parts are independent and don't have a built-in way to remember what happened in the previous steps of a sequence.

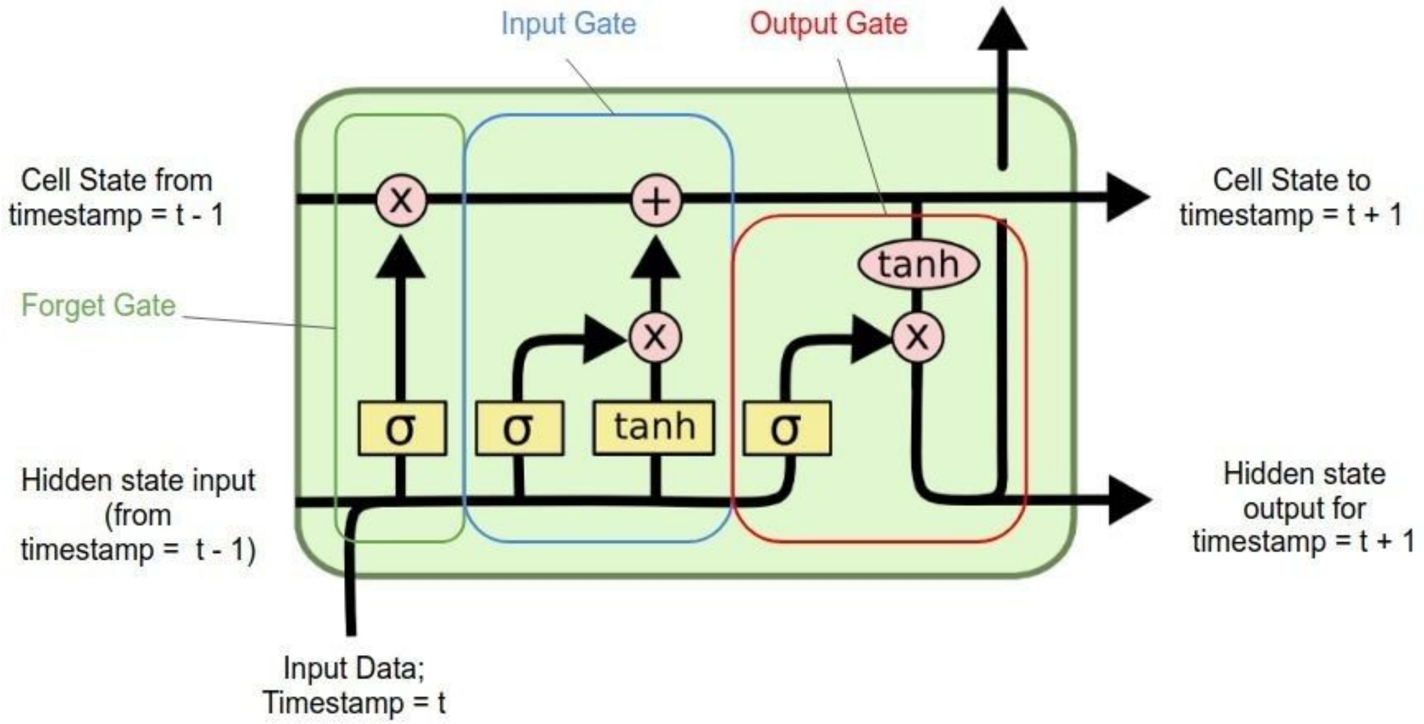


Recurrent Neural Network



Feed-Forward Neural Network

Recurrent Neural Networks: ***LSTM***



Recurrent Neural Networks: ***Task***

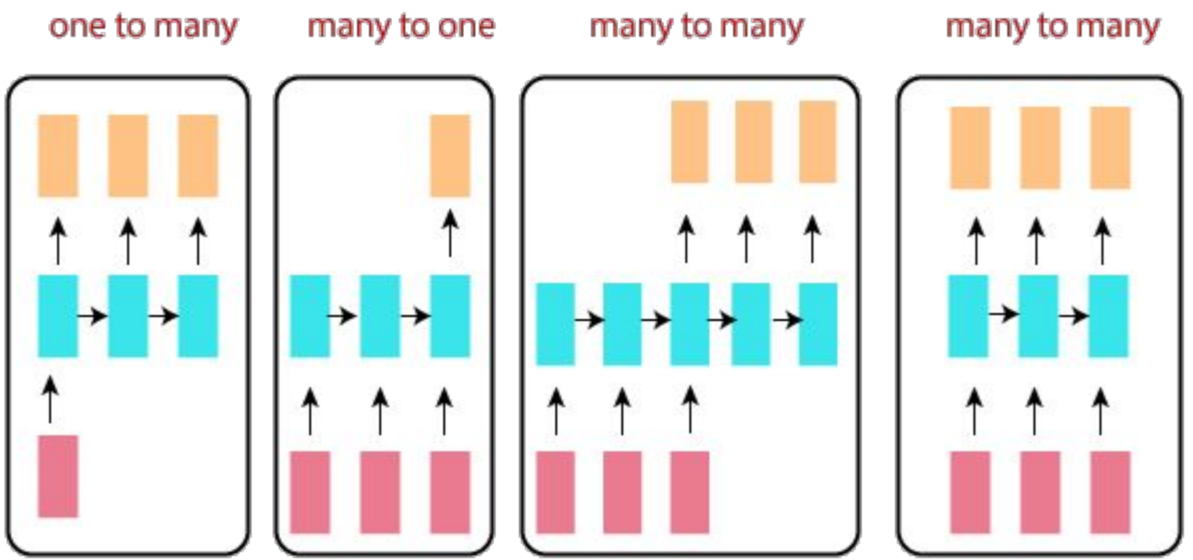


Image captioning

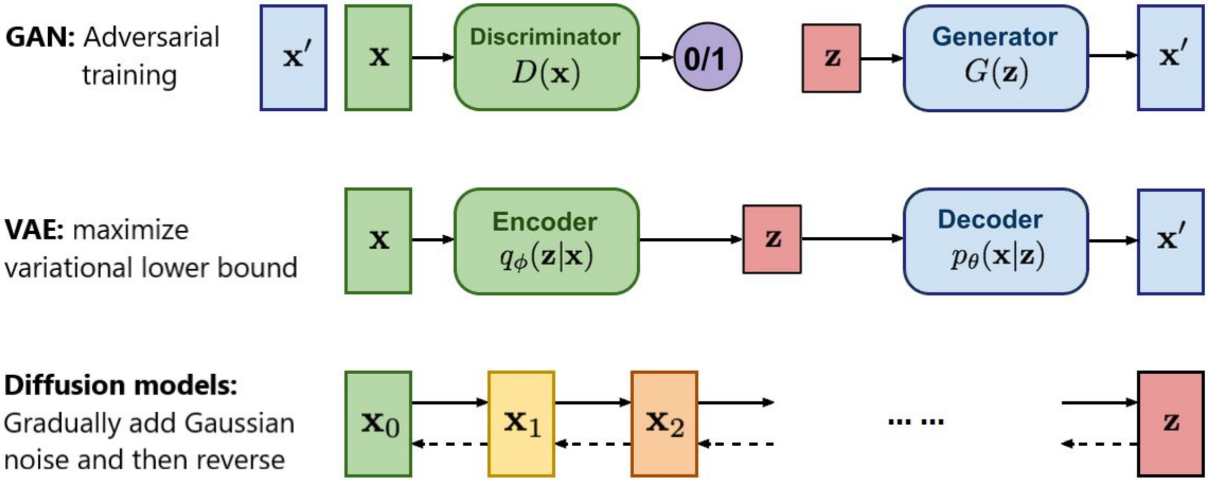
Sentiment analysis

Text generation or **machine translation**

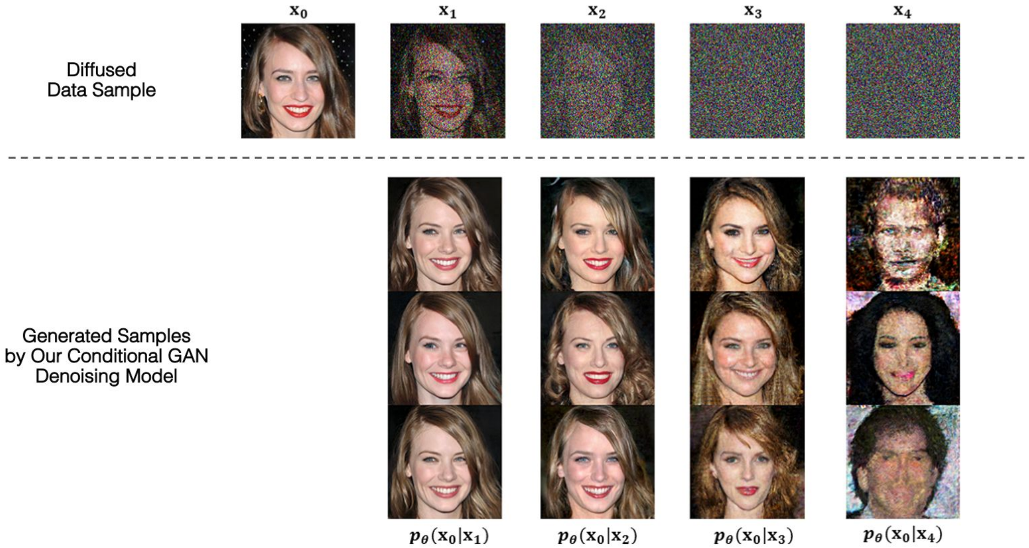
Speech tagging or **named entity recognition**

Generative Neural Network (GANs)

A class of models designed to generate new data that is similar to the training data they've been fed. They learn the underlying distribution of a dataset and then use this knowledge to produce new instances that could plausibly come from the same distribution.



Generative Neural Networks: ***Task***



The website *This Person Does Not Exist* was created in February 2019. It uses GANS, to generate highly realistic images of human faces of people who do not actually exist.
https://this-person-does-not-exist.com/en

04.

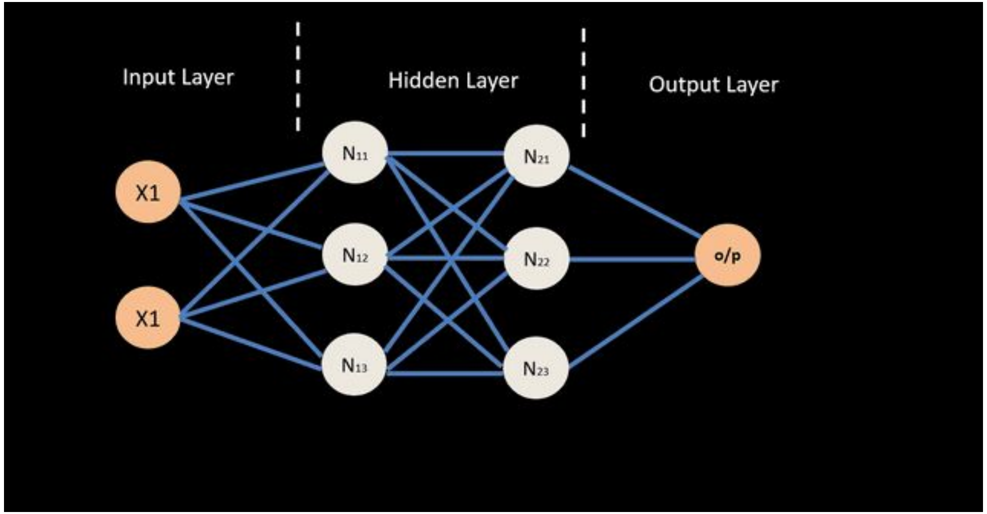
Training

Here, we discuss the process of training machine learning models, including data preparation, model selection, and the use of algorithms to optimize model performance. We'll also cover strategies to avoid common pitfalls like overfitting.

Training

Backpropagation

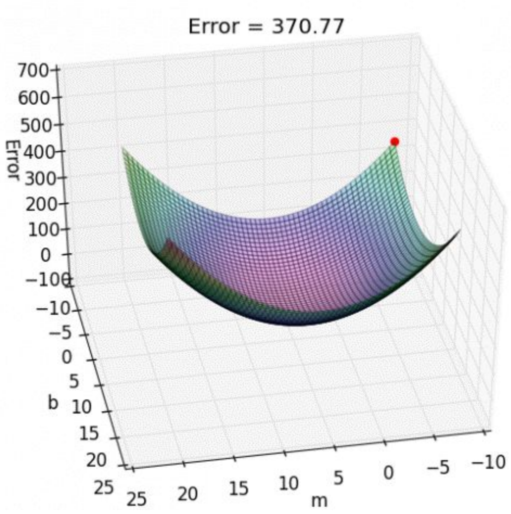
An algorithm used to calculate the gradient of the loss function with respect to each weight by the chain rule, efficiently propagating the error backward through the network.

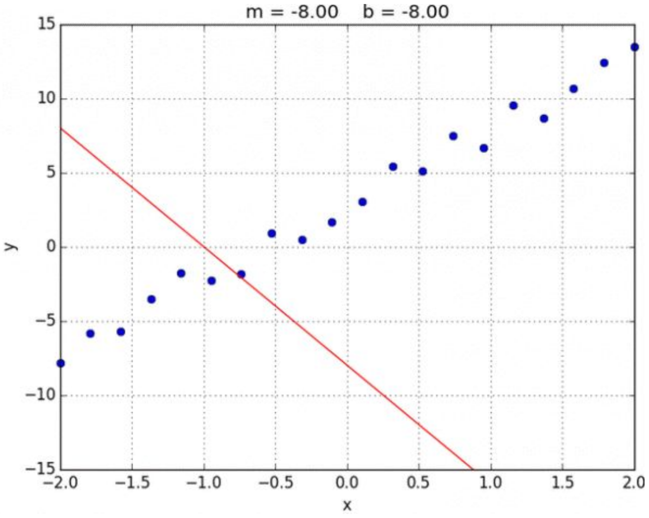


Training

Backpropagation

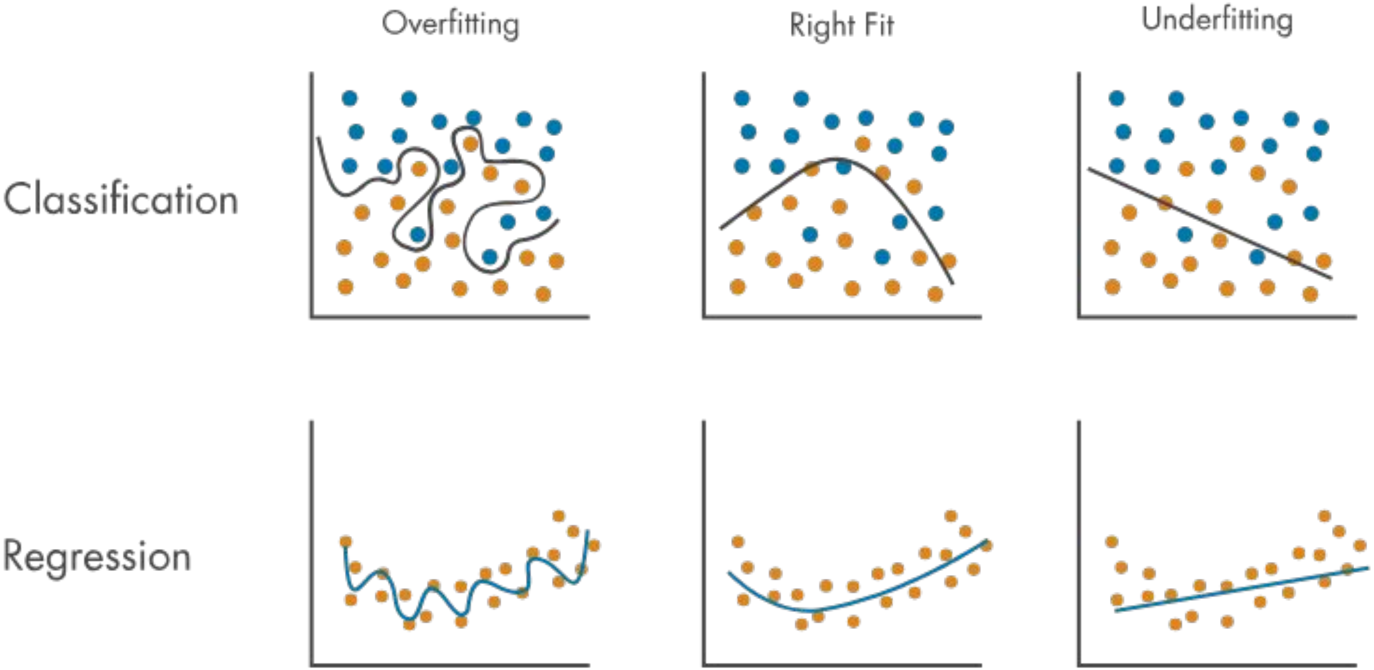
An algorithm used to calculate the gradient of the loss function with respect to each weight by the chain rule, efficiently propagating the error backward through the network.





Training

Overfitting and underfitting



Generalization

Model's ability to perform well on new, unseen data after being trained on a specific dataset. It measures the effectiveness in applying learned patterns to novel inputs outside the training set.

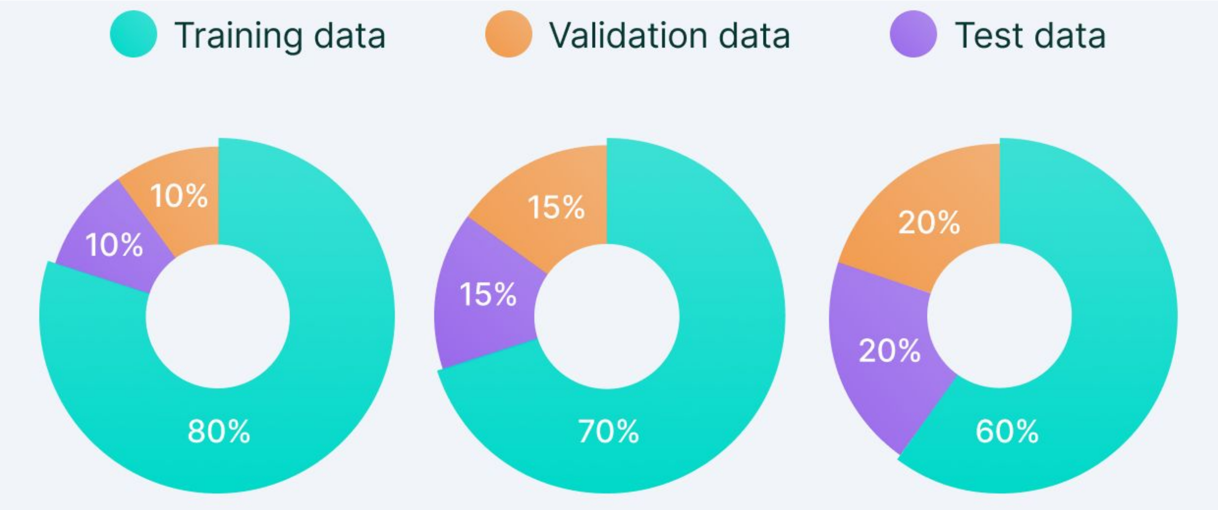
Training data

Generalization

Testing data

Generalization: ***Set splitting***

A process a dataset is divided into separate subsets to ensure that models are trained on one set of data and tested on unseen data to evaluate performance and generalize ability.



Generalization: ***Data augmentation***

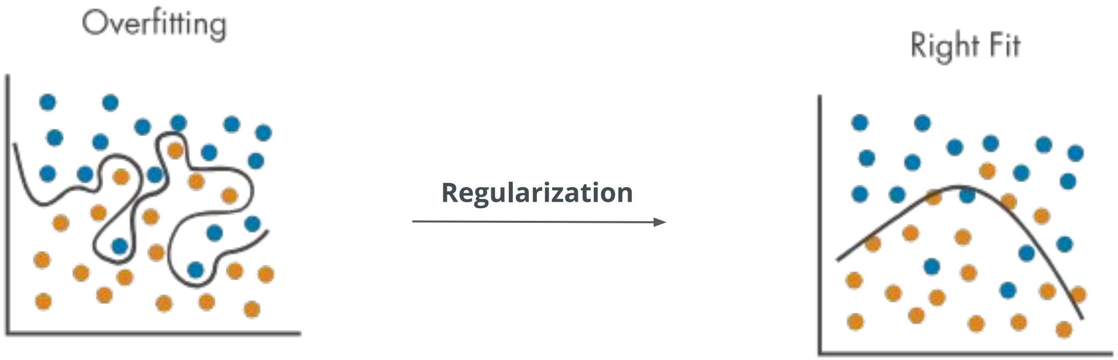
A technique to increase the diversity of training data by applying various transformations, such as rotation, scaling, and flipping, to existing samples. This helps improve model robustness and generalization by simulating a wider range of input scenarios.

Original

Augmentation

Generalization: ***Regularization***

A technique used to prevent overfitting by adding a penalty on the size of the parameters. It encourages simpler models during training, which can generalize better on unseen data.



05.

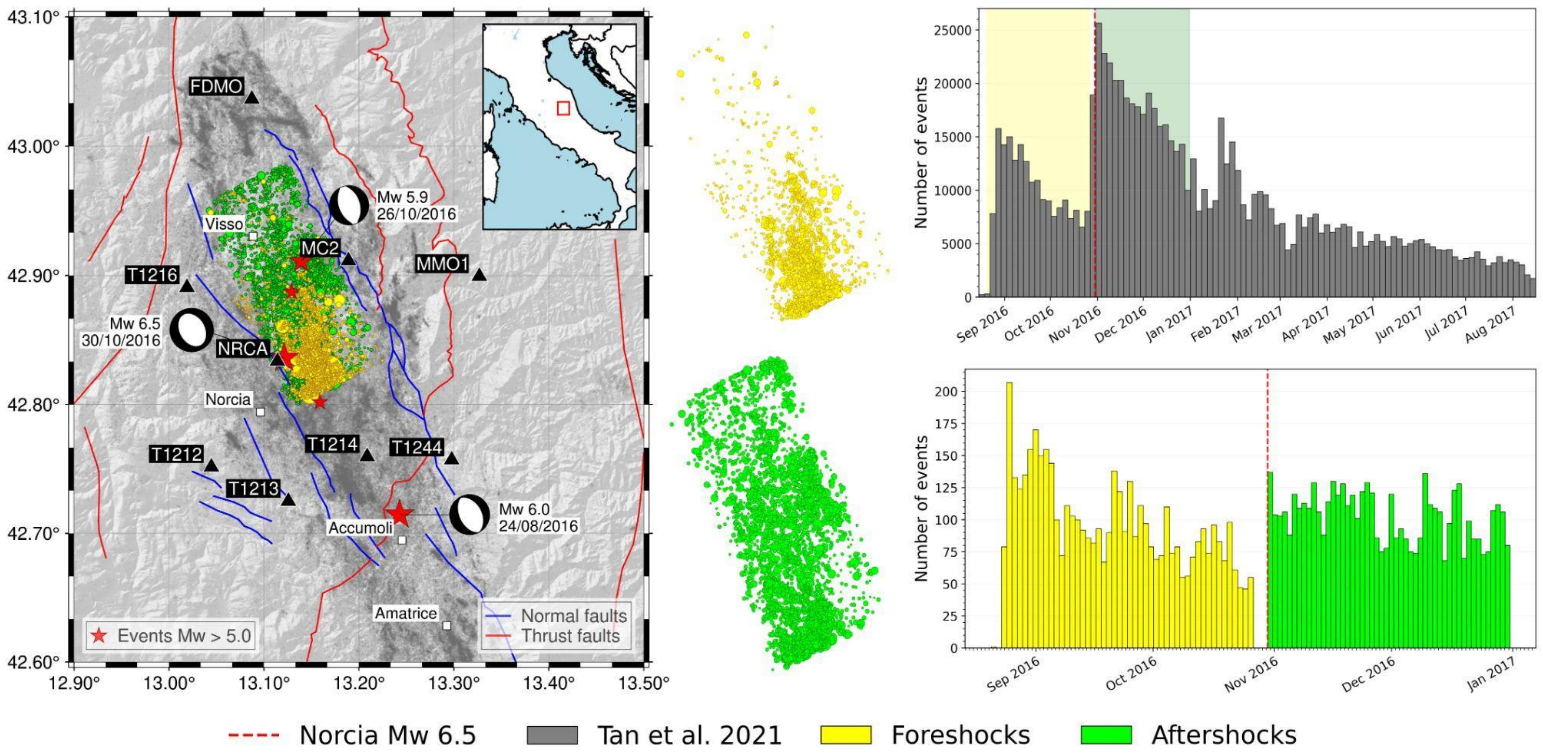
Applications

This section delves into the wide range of applications for machine learning and deep learning in seismology.

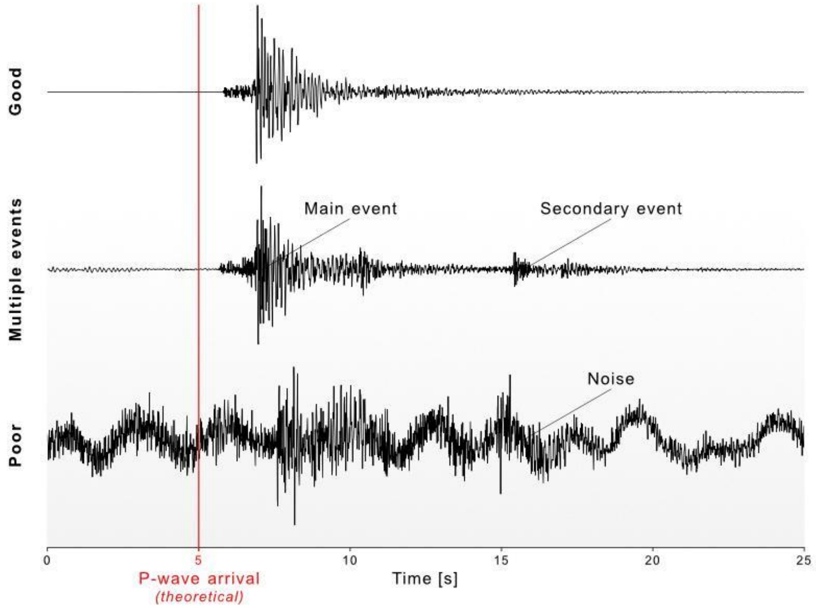
05.01

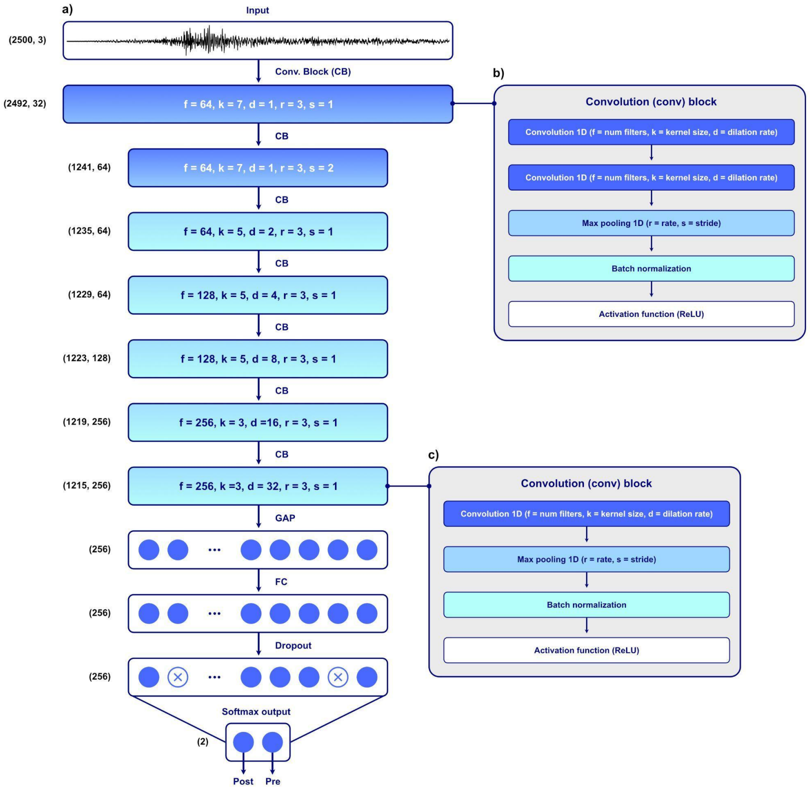
Earthquake classification

Earthquake classification



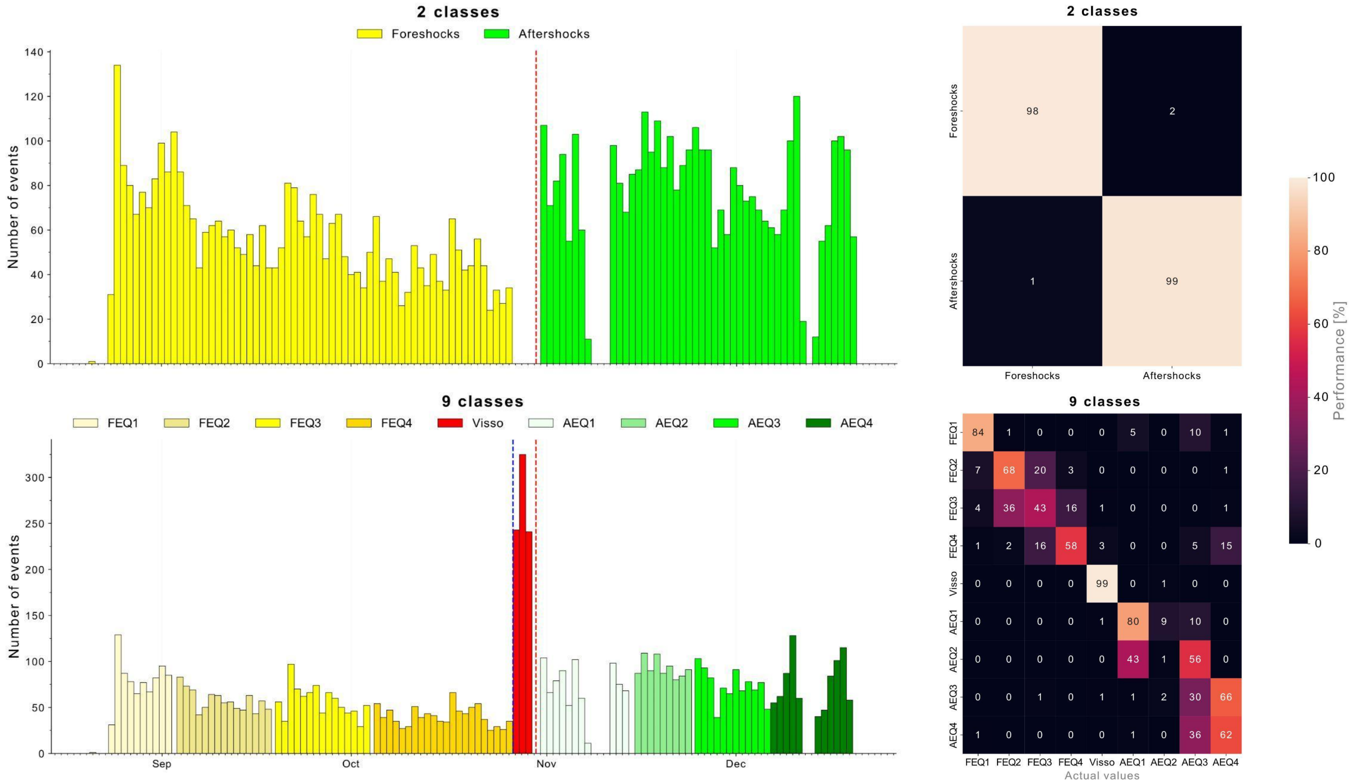
Earthquake classification: ***Methods***





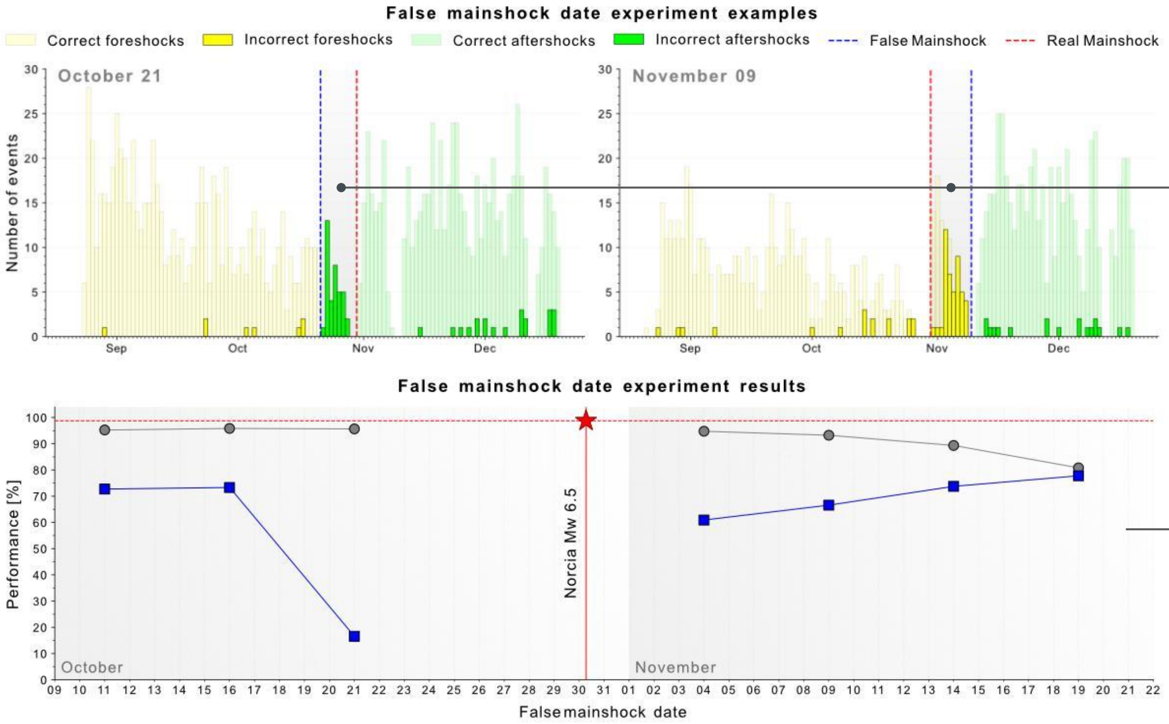
Applications

Earthquake classification: ***Results***



Applications

Earthquake classification: ***Validation***



Re-labeled data

Results obtained after re-labeling

05.02

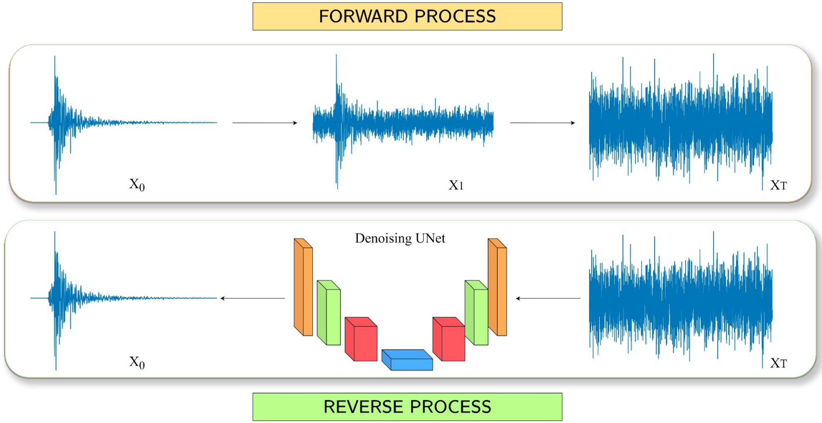
Cold Diffusion model

Applications

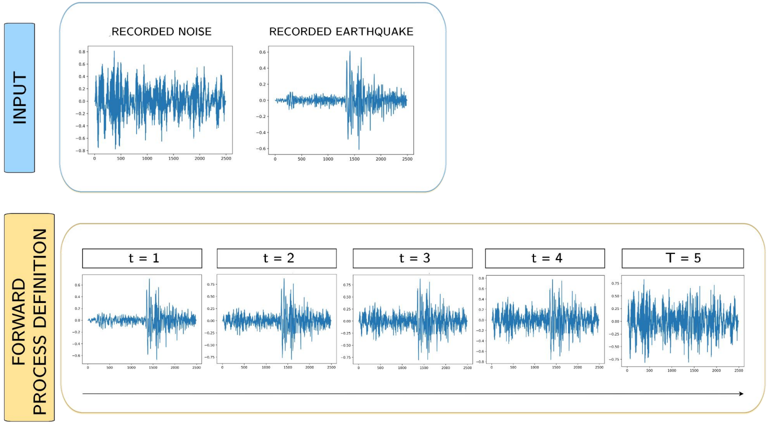
Cold Diffusion model: ***Methods***

Our research explores adapting the Cold Diffusion model for seismic denoising, tailoring it to overcome the challenge of non-Gaussian noise in seismic data, promising enhanced signal recovery.

Diffusion Process

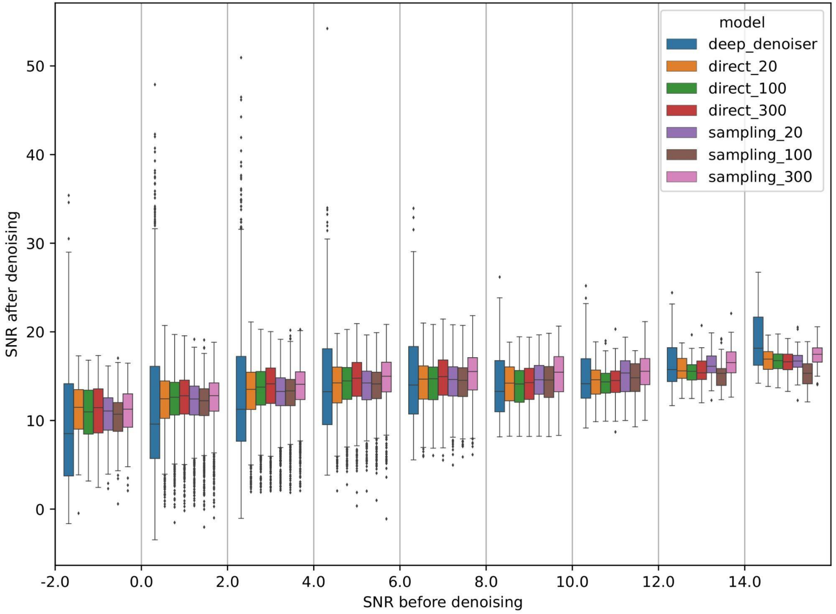


Input Assumptions

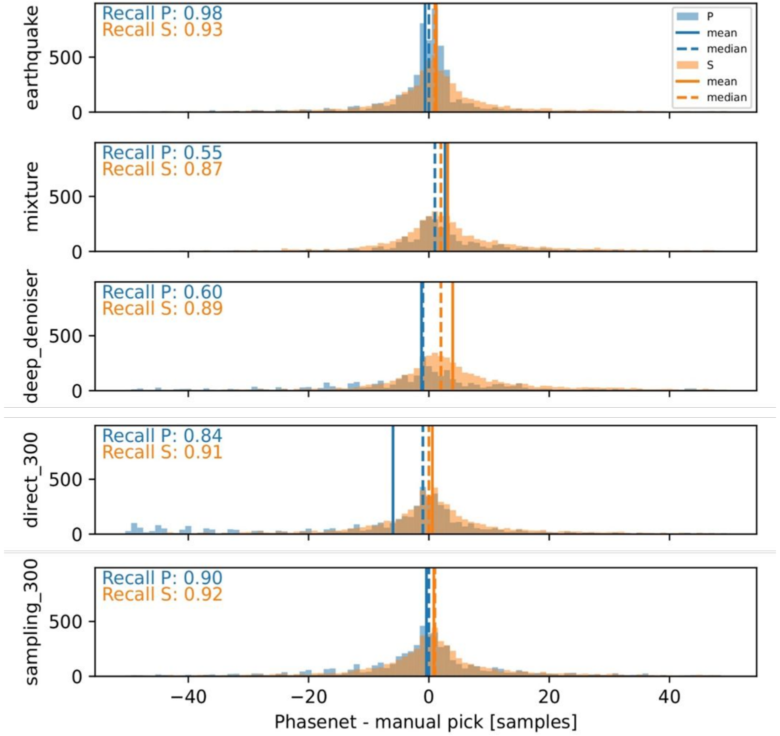


Cold Diffusion model: ***Quantitative results***

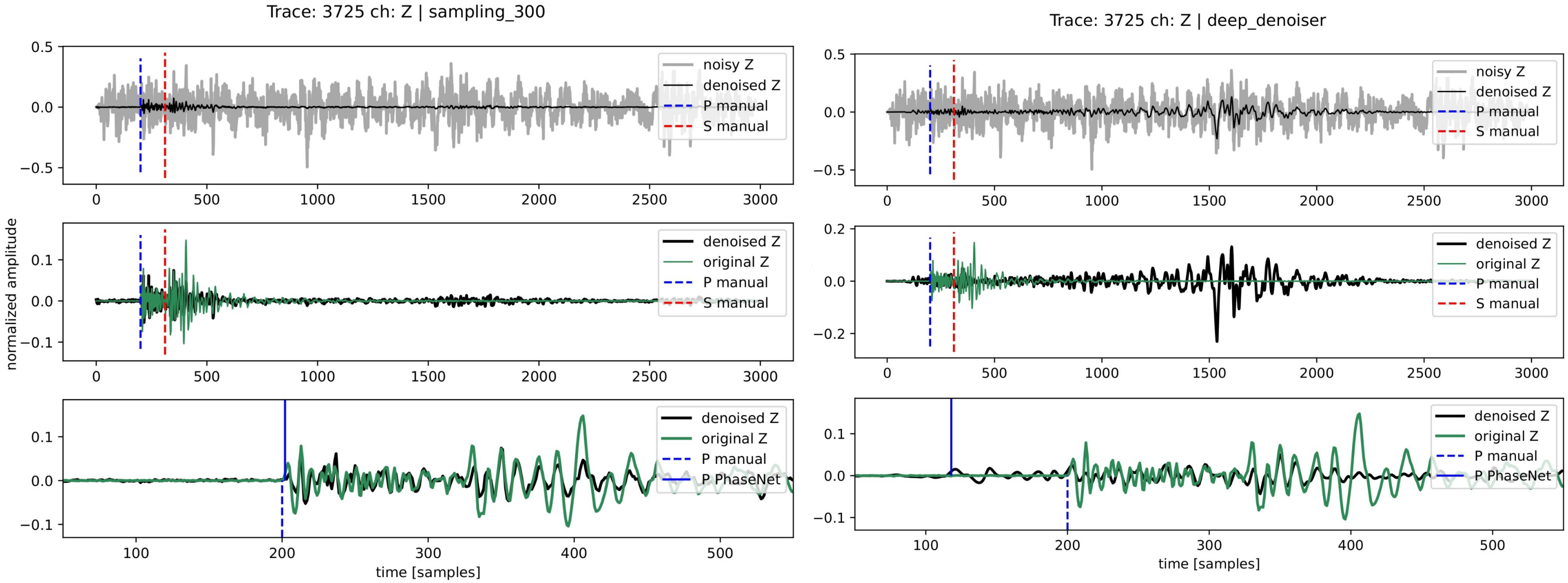
Distributions of SNR values before and after the denoising



Phase arrivals picks



Cold Diffusion model: ***Qualitative results***



Elements of Seismology & Machine Learning

Now it's your turn to apply machine learning to seismology!
If you have any ideas or questions, here are our contact details:

Daniele Trappolini
daniele.trappolini@uniroma1.it

Giacomo Mastella
giacomo.mastella@uniroma1.it