



ITINERIS



INTRODUCTION TO MACHINE LEARNING

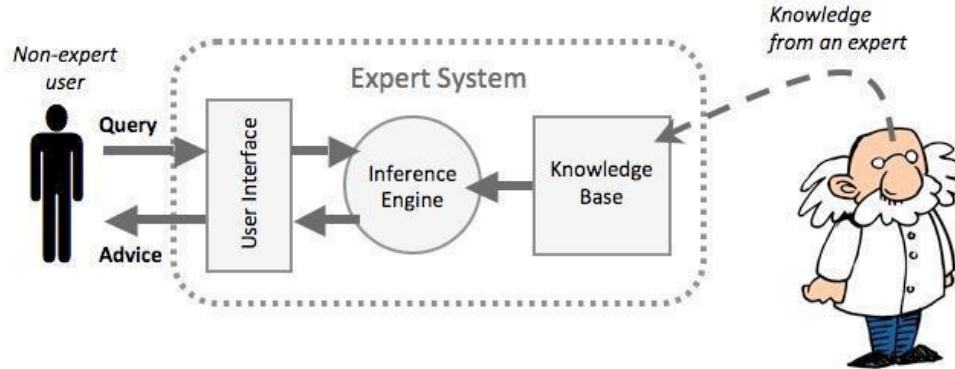
Prof. Alessia Saggese

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 "Education and Research" - Component 2: "From research to business" - Investment
3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"



WEAK THESIS AND EXPERT SYSTEMS

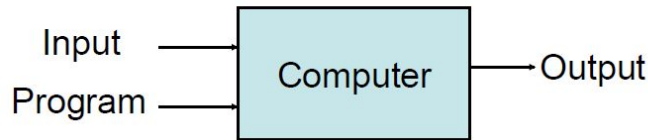
The weak thesis attempts to reproduce human capacities, using reasoning engines (inferential) that start from a knowledge base consisting of assertions, rules and relationships and generate new knowledge through induction and deduction.



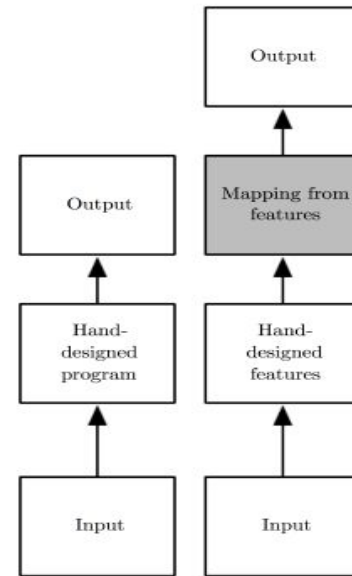
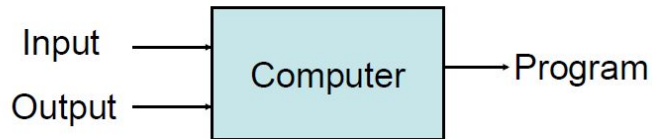
STRONG OR CONNECTIONIST THESIS

The discipline of Machine Learning studies algorithms capable of learning to perform a particular task using experience, i.e. learning through examples.

Traditional Programming



Machine Learning



- ▶ To solve a problem, you need to know a solution algorithm.

Example of the second degree equation $ax^2+bx+c=0$

Calculate $\Delta=b^2-4ac$

If $\Delta < 0$ then there are no real solutions

If $\Delta = 0$ there are two coincident solutions $x_1=x_2=(-b)/2a$

If $\Delta > 0$ there are two distinct real solutions:

$$x_1=(-b-\sqrt{\Delta})/2a \text{ and } x_2=(-b+\sqrt{\Delta})/2a$$

How to solve the problem without knowing the algorithm?

We use learning from examples (Machine Learning)

► Training-set:

$A=1, B=-3, C=2$ $x_1=1, x_2=2$

$A=1, B=-10, C=25$ $x_1=x_2=5$

$A=1, B=0, C=25$ no real solution

..... And thousands of other examples whose solution is known

The machine learns to solve unseen equations starting from the examples presented in the training set

The more examples I provide, the better the machine will be able to respond!

Now she is ready to answer...

$A=1.02, B=-10.01, C=23.99$ the machine responds $x_1=4.16$ $x_2=5.65$

STRONG AI APPROACHES

- ▶ Handcrafted Features

Historical approach, requires you to have industry experts in the design phase of the system

Requires (relatively) little data

Provides more explainable decisions

Deep Learning

Doesn't require (a lot) of industry experience on the part of the designer

Requires a large amount of data

The reasons for the decision may not be explicit

EXAMPLE: CHECK FOR PPE

- ▶ As a reference example to describe the differences between the handcrafted approach and the deep approach, consider the problem of verifying the presence of personal protective equipment (PPE) worn by workers.

In particular, we focus on the 4 classes typically required in real products: helmet, mask, vest and gloves.



Helmet



Mask

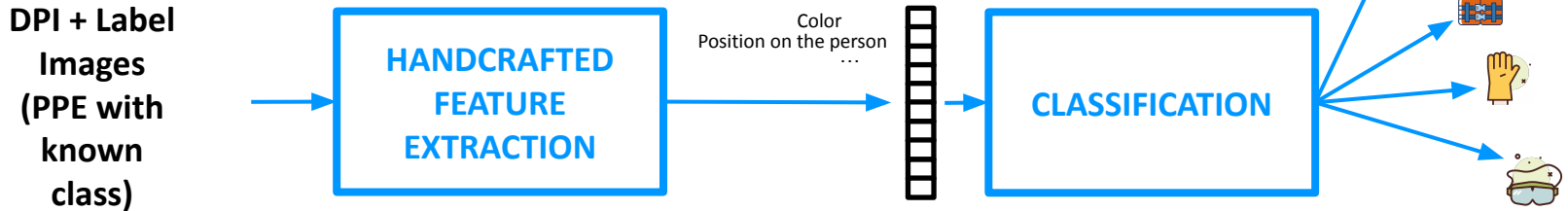


Waistcoat



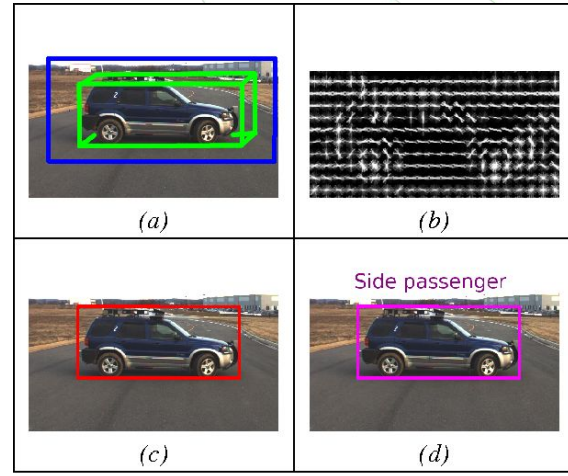
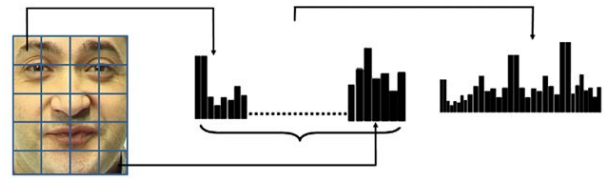
Gloves

- ▶ The AI application designer identifies the features (e.g. PPE color, position on the person, among others) that allow the specific personal protective equipment to be identified. During the learning phase, the machine represents examples through these features and learns to classify the input images by recognizing the DPI. What if there were better features?



FEATURES E DATA REPRESENTATION

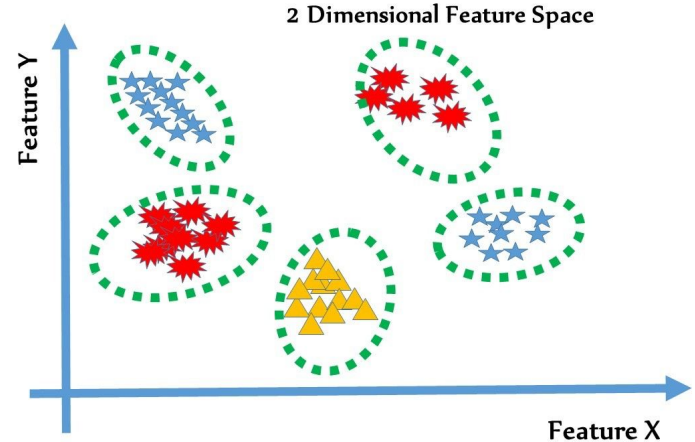
The examples used are represented by features that allow you to maximize the differences between different objects and minimize the variations between similar elements.



VECTOR REPRESENTATION OF OBJECTS

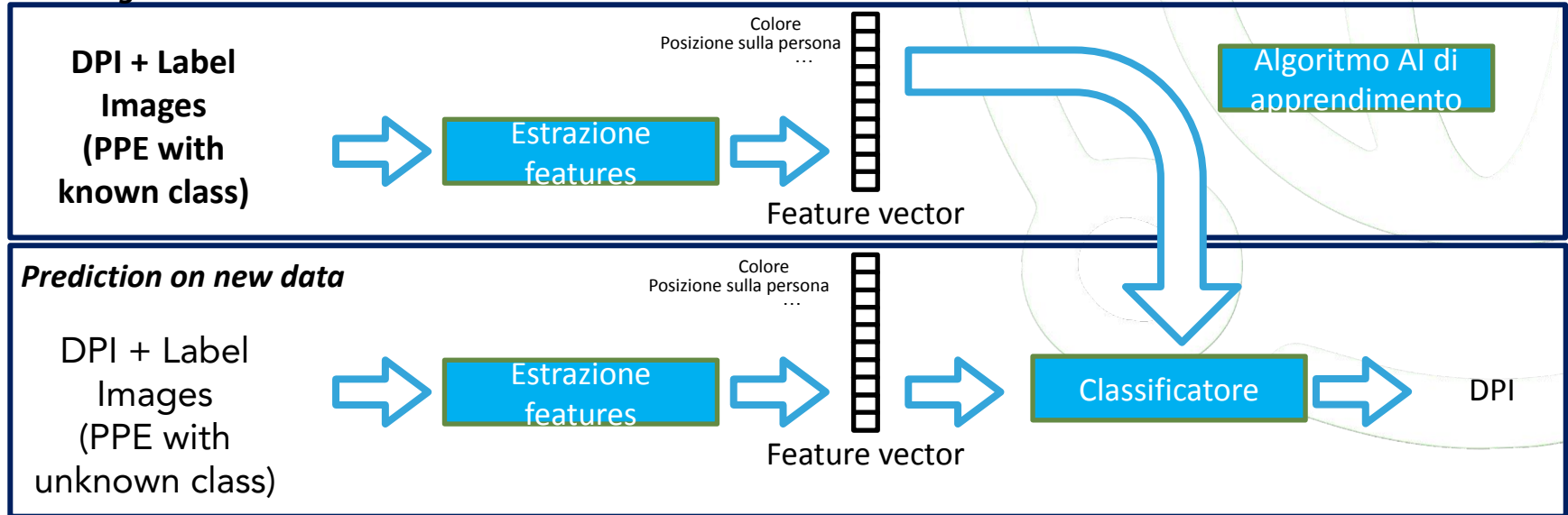
- Once the features are measured, a vector (Characteristic Vector) is obtained

Each particular object corresponds to a vector
If the Features are n , we consider an n -dimensional space and the object is a point in space



THE HANDCRAFTED APPROACH

Training



- Learning from examples means in this case:
Knowing how to choose the right features
To be able to build, starting from examples, the best decision-making regions

REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Helmet



Jacket



Gloves

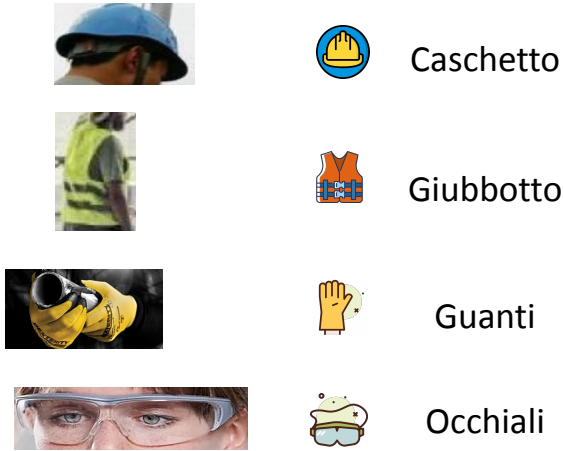


Glasses

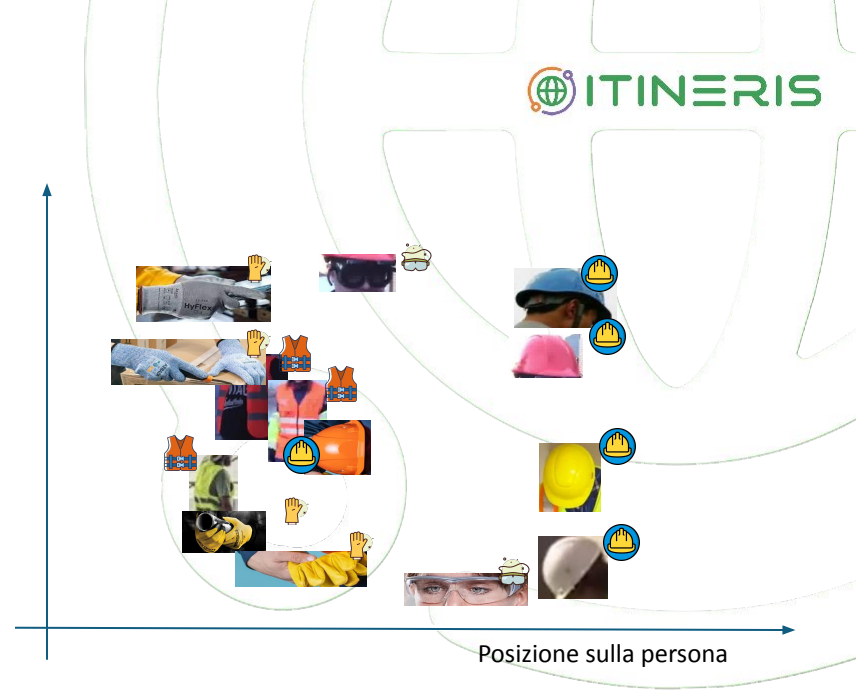
- ▶ Examples used are represented by features.
Once the features of each sample have been measured, a Features Vector is obtained
Each sample can then be represented as a vector
If the features are "n" we consider an "n"-dimensional space and the sample becomes is a point in vector space

REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label

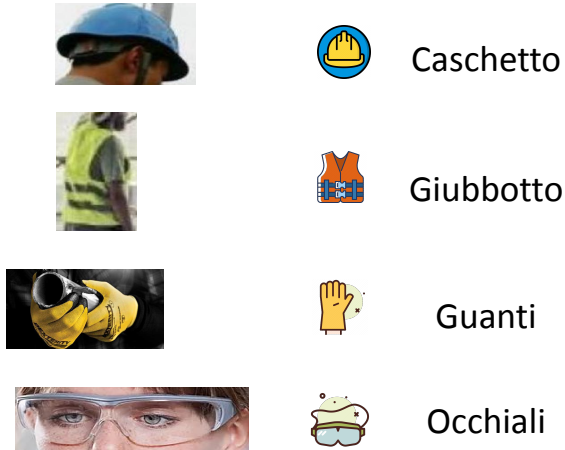


Colore



REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Colore



Features must maximize differences between different DPIS and minimize variations between images belonging to the same DPI.

REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Caschetto



Giubbotto



Guanti



Occhiali

- ▶ During the training phase, the goal is: Choose the most appropriate features (such as to maximize differences between different patients and minimize variations between similar patients)
Build, starting from examples, the best decision-making regions

REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Caschetto



Giubbotto



Guanti



Occhiali



REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Caschetto



Giubbotto



Guanti



Occhiali



REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Caschetto



Giubbotto



Guanti



Occhiali



REPRESENTATION OF THE PROBLEM

Input: immagine con DPI + Label



Caschetto



Giubbotto



Guanti



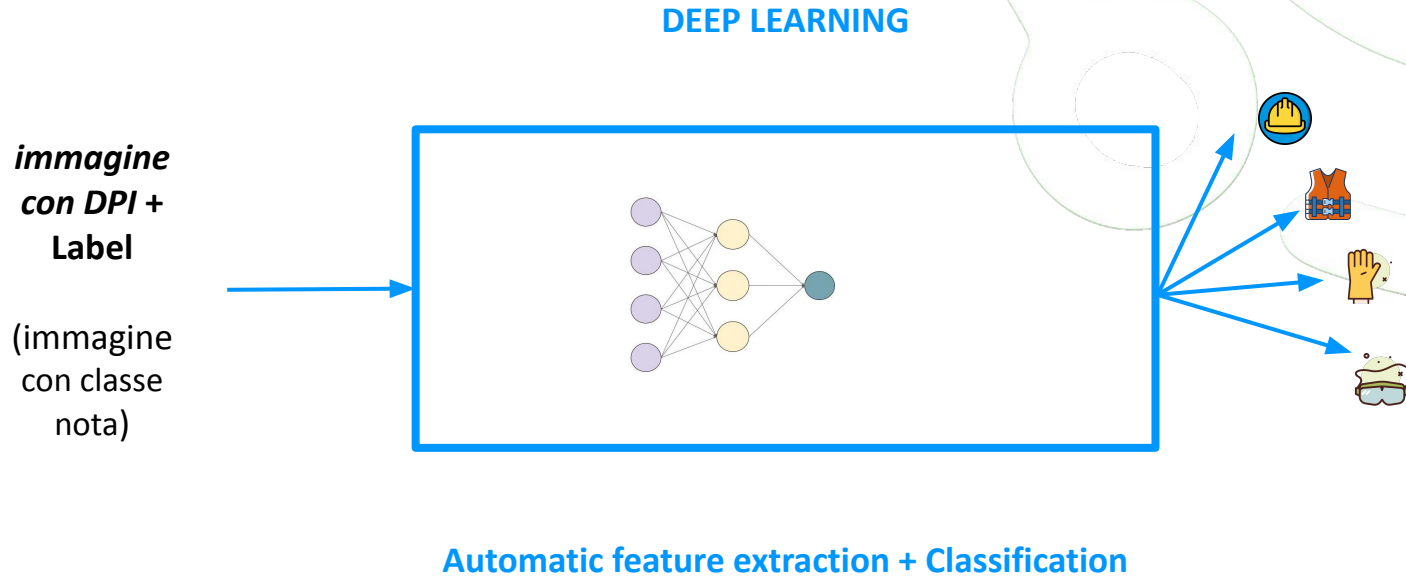
Occhiali



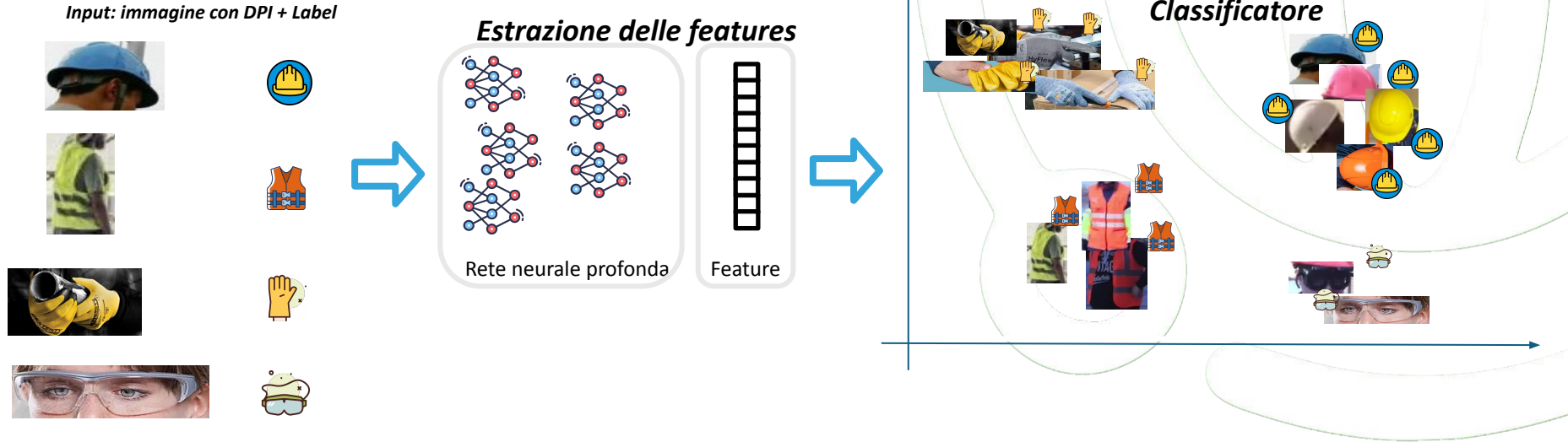
They are
good
features!

THE DEEP LEARNING APPROACH

- ▶ The machine learns autonomously with examples to understand which features allow it to recognize the specific PPE.

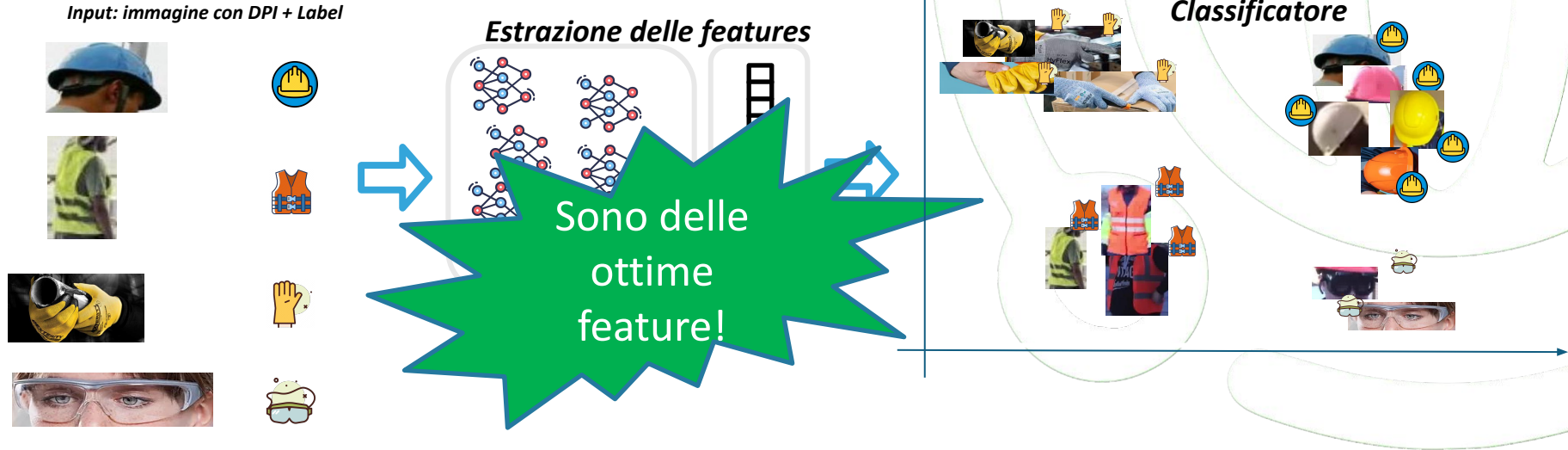


THE DEEP LEARNING APPROACH



- ▶ The best features are learned directly from the data (images with label notes) and are much more discriminating; in fact, the categories of PPE are much further apart.

THE DEEP LEARNING APPROACH



- ▶ The best features are learned directly from the data (images with label notes) and are much more discriminating; in fact, the categories of PPE are much further apart.

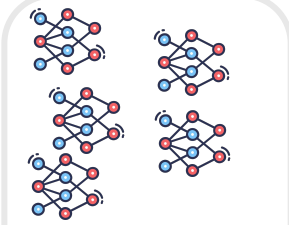
THE DEEP LEARNING APPROACH

Input: immagine con DPI + Label

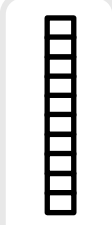
Input: una nuova immagine



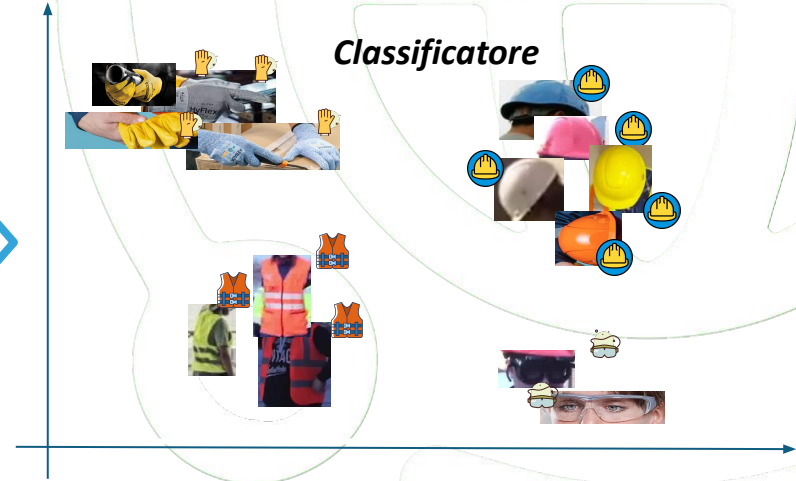
Estrazione delle features



Rete neurale profonda



Feature



- ▶ **Classification:** given the feature vector, the system determines the expected class to which the image belongs on the basis of the decision regions

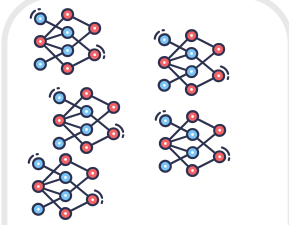
THE DEEP LEARNING APPROACH

Input: immagine con DPI + Label

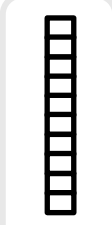
Input: una nuova immagine



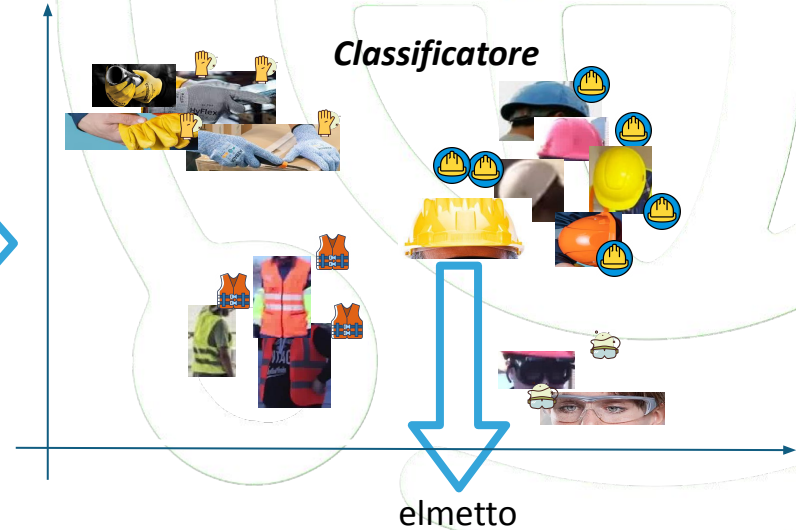
Estrazione delle features



Rete neurale profonda



Feature



- ▶ **Classification:** given the feature vector, the system determines the expected class to which the image belongs on the basis of the decision regions

WHY AI IS SO POWERFUL

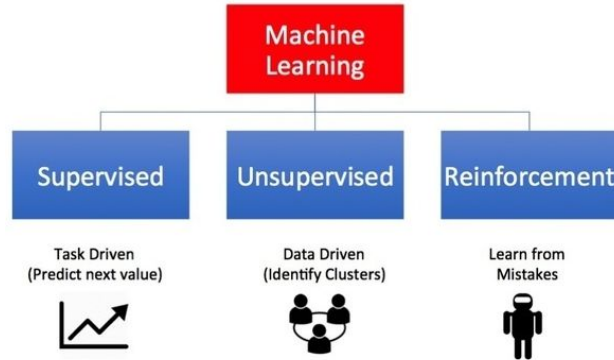
- ▶ Because the number of examples can grow enormously, increasing the accuracy of the model learned
It can outperform human performance when TS is large and representative
For example, ChatGPT is trained with a million million words
A human being reads 250 words per minute; it takes 22,000 years JUST to read the same number of words.



MODELS FOR LEARNING

LEARNING METHODS

- ▶ **Supervised Learning:** the intelligent system learns from examples of the problem under consideration whose solution is known.
- Unsupervised Learning:** the intelligent system learns to divide data into homogeneous groups (clusters) with respect to the application objective
- Reinforcement Learning:** The intelligent system is pre-trained to perform the task, but then improves over time by learning from mistakes



SUPERVISED LEARNING

Nell'apprendimento supervisionato, i dati di addestramento sono coppie, ognuna costituita da un input X e dal relativo output (label) Y

Task

- ▶ Si vuole prevedere una funzione sconosciuta $Y = F(X)$, dati X e Y

Esperienza

- ▶ Si riceve un insieme di coppie $(X_1, Y_1), (X_2, Y_1), \dots, (X_n, Y_n)$

Performance

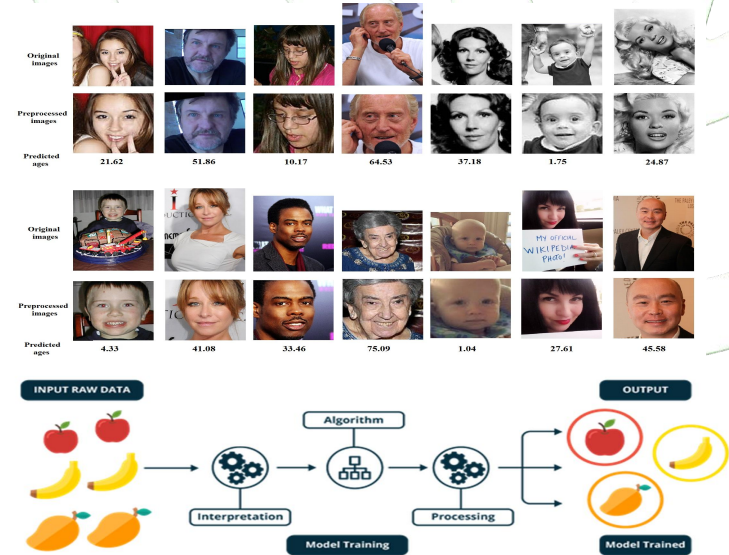
- ▶ Una metrica capace di misurare l'errore che si commette nella predizione di $F(X)$

SUPERVISED LEARNING

Supervised learning problems are divided into two types:

Regression: the output to be recognized has continuous values (estimate of the price of a house, estimate of the age of a person, financial analysis, etc.)

Classification: the output to be recognized is a class or category (recognition of a person's gender, sentiment analysis, classification of tumors, etc.)



SUPERVISED LEARNING

In supervised learning, training data is pairs, each consisting of an input and its output (label)

X_1 :



$F(X_1)$:

SUPERVISED LEARNING

In supervised learning, training data is pairs, each consisting of an input and its output (label)

X_2 :



$F(X_2)$:

Gatto

SUPERVISED LEARNING

In supervised learning, training data is pairs, each consisting of an input and its output (label)

X_3 :



$F(X_3)$:

Cane

SUPERVISED LEARNING

In supervised learning, training data is pairs, each consisting of an input and its output (label)

X_n :



$F(X_n)$:

Gatto

SUPERVISED LEARNING

X :



$F(X)$:

?

UNSUPERVISED LEARNING

- ▶ Nell'unsupervised learning, i dati di addestramento includono unicamente un set di input, non le relative label

Task

- ▶ Si desidera suddividere i dati di input in gruppi omogenei rispetto all'applicazione.

Esperienza

- ▶ L'algoritmo di apprendimento riceve un insieme di dati di input X_1, X_2, \dots, X_n

UNSUPERVISED LEARNING

- Example

A recommendation system (e.g. Amazon):
Must divide users into groups (unsupervised learning)
Must recommend goods purchased by other users in the same group to a user

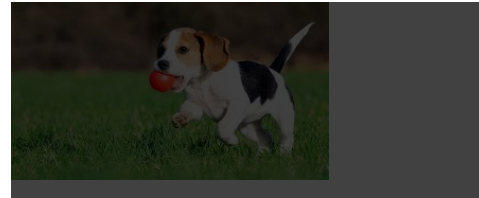
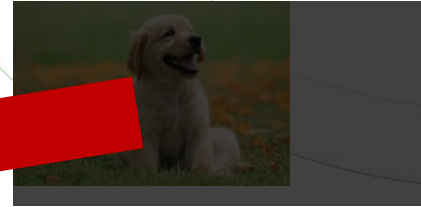
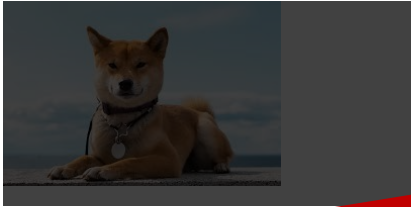
UNSUPERVISED LEARNING

Suppose we want to simulate an unsupervised learning algorithm that generates two clusters from the following images:



UNSUPERVISED LEARNING

Suppose we want to simulate an unsupervised learning algorithm that generates two clusters from the following images:



This is the possible cluster 1

UNSUPERVISED LEARNING

Suppose we want to simulate an unsupervised learning algorithm that generates two clusters from the following images:



This is possible cluster 2

REINFORCEMENT LEARNING

- ▶ Nel reinforcement learning, i dati di addestramento includono unicamente un set di input

Task

- ▶ Si vuole prevedere una funzione sconosciuta $F(X)$ data X ; la scelta di $F(X)$ influenzerà la prossima X che sarà inviata al sistema (si pensi a $F(X)$ come a un'azione).

Esperienza

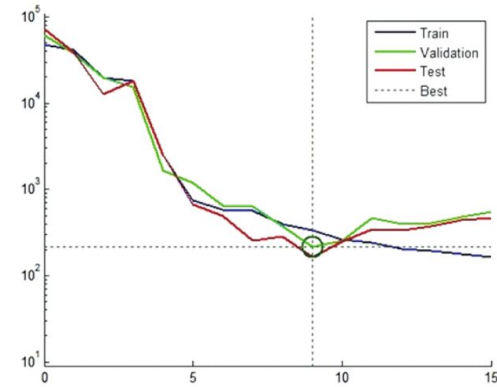
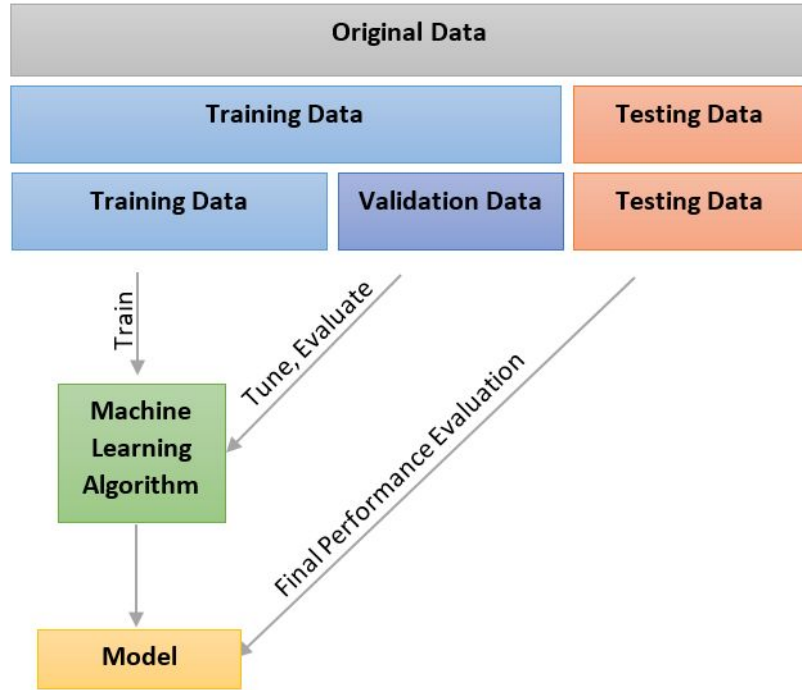
- ▶ Si possono eseguire diversi esperimenti, ottenendo la sequenza X_1, X_2, \dots, X_n ed uno score di performance legato a tale sequenza $P(X_1, X_2, \dots, X_n)$

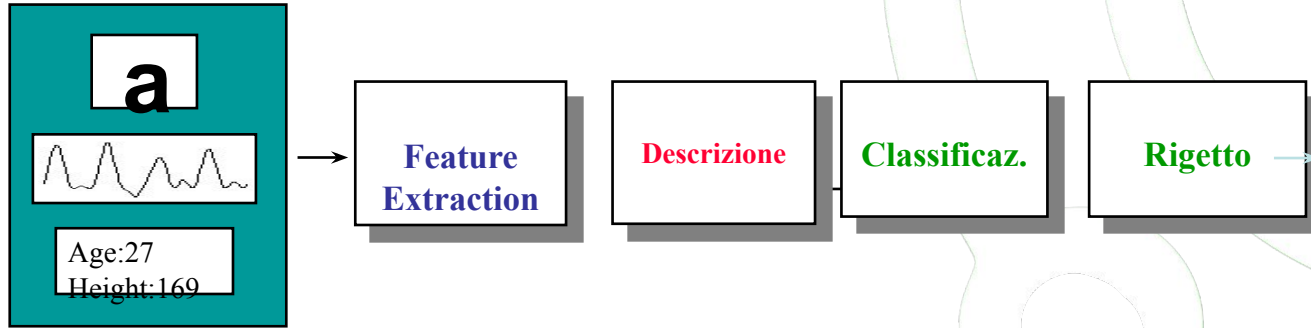
REINFORCEMENT LEARNING

- Example
Each of us has experienced reinforced learning firsthand:
 $F(X)$: the action on the pedals and handlebars
 $P(X_1, \dots, X_n) = -1000$ if you lose your balance, $+10$ if you follow the predicted trajectory



TRAINING, VALIDATION E TEST





Raw data

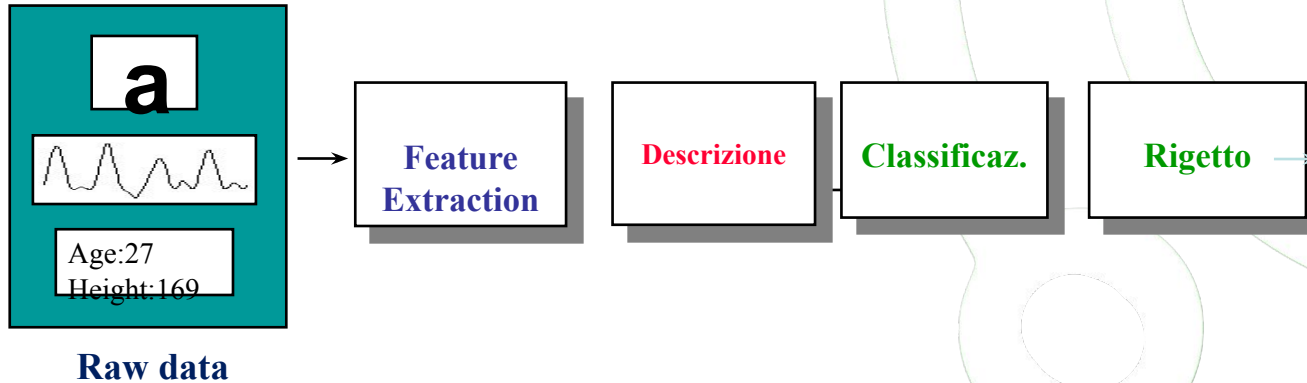
Features:

They represent objects of interest in the smart system

They must unite all the champions of the same class

They must be "Discriminating" with respect to samples of different classes

They must be in adequate numbers



Vector Representation of Objects

Once the features are measured, a vector (Characteristic Vector) is obtained

Each particular object corresponds to a vector

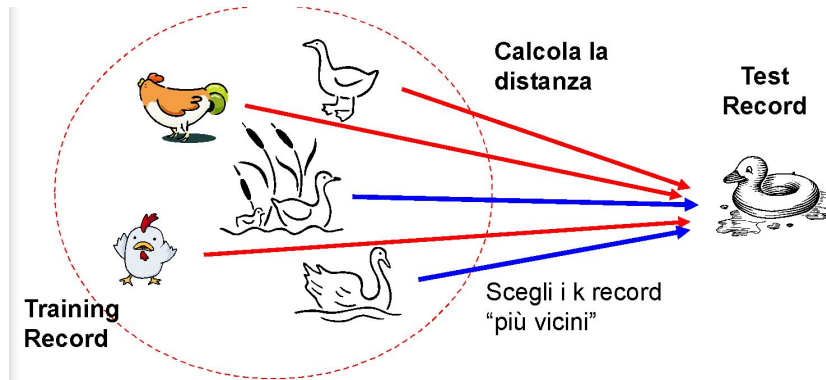
If the Features are n , we consider an n -dimensional space and the object is a point in space

CLASSIFICATORI NEAREST NEIGHBOR (NN)

- ▶ They use the k "nearest" points (nearest neighbors) to make the classification

Basic idea:

If it walks like a duck, quacks like a duck, then it's probably a duck



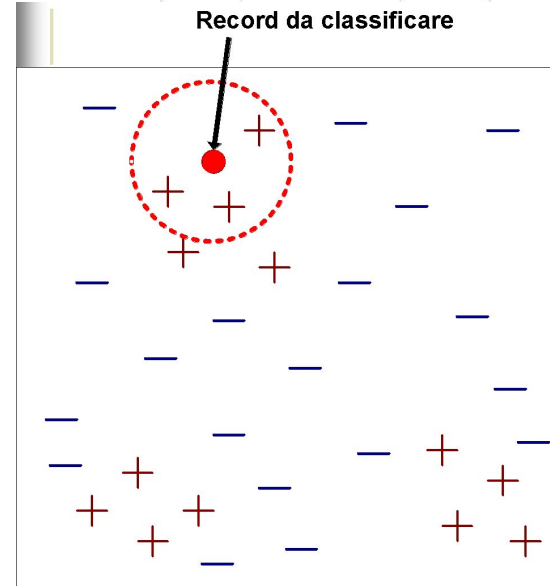
NEAREST NEIGHBOR CLASSIFIERS (NN)

Require:

A training set

A metric to calculate the distance between records

The value of k , i.e. the number of neighbors to be used



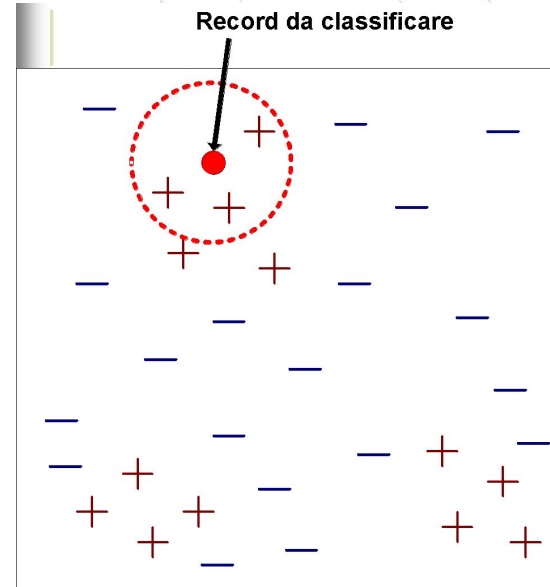
NEAREST NEIGHBOR CLASSIFIERS (NN)

The classification process:

Calculate distance to records in the training set

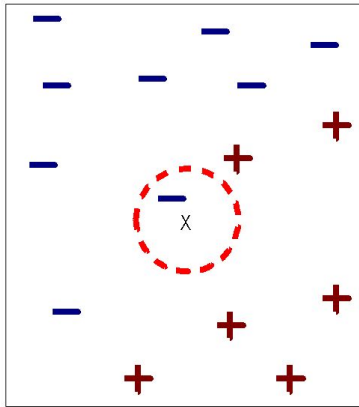
Identify k nearest neighbors

Use the class labels of the nearest neighbors to determine the class of the unknown record (e.g. choosing the one that appears most frequently)

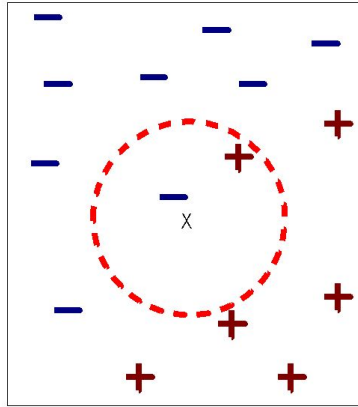


NEAREST NEIGHBOR CLASSIFIERS (NN)

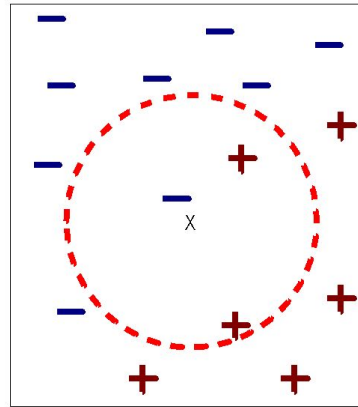
- ▶ The k nearest neighbors of a record x are the records in the training set that have the smallest k distances from x



(a) 1-nearest neighbor



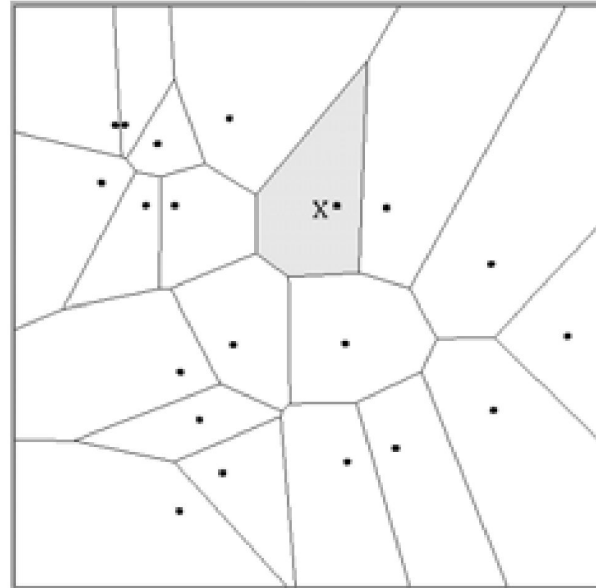
(b) 2-nearest neighbor



(c) 3-nearest neighbor

1-NEAREST NEIGHBOR (NN)

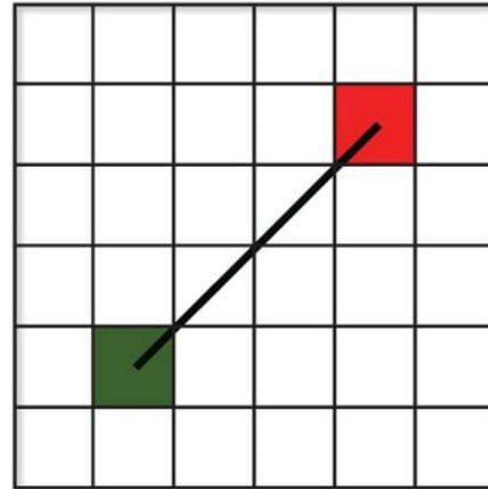
- ▶ Form a Voronoi tessellation



DISTANCE METRIC

- ▶ How do I calculate distance?
One of the widely used metrics is Euclidean distance:

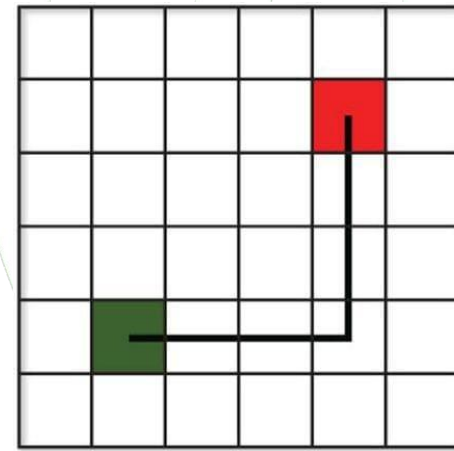
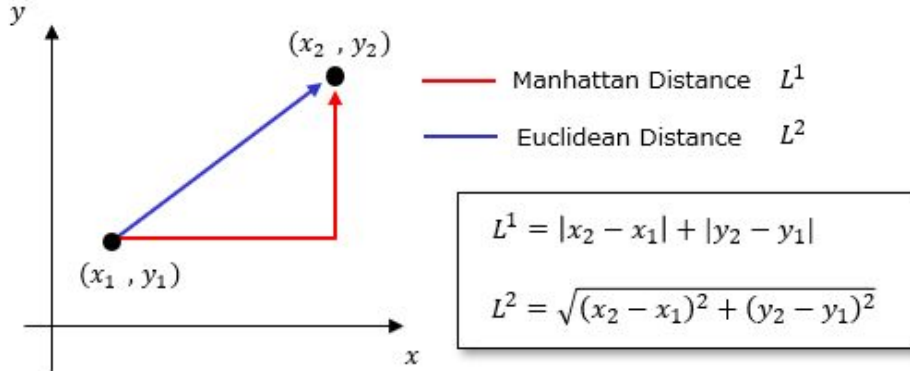
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$



Euclidean Distance

DISTANCE METRIC

- ▶ Of course, it's not the only one...
Ex: Distance Manhattan



Manhattan Distance

ITINERIS

K-NEAREST NEIGHBOR

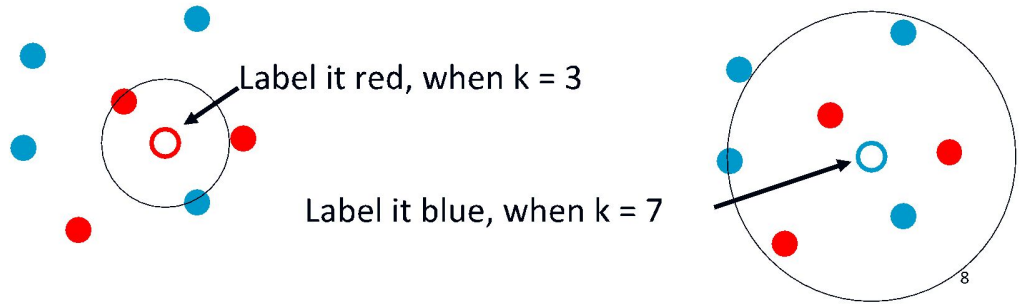
- ▶ The classification of a z record is obtained by a process of majority voting among the k elements D_z of the training set D closest to (or similar) to z

Note: This way, all neighbors have the same weight. We could decide to weigh the contribution of neighbors based on distance

Ex: $w = 1/d(z, x_i)$

COME SCEGLIERE K?

- ▶ The choice of k is important because:
 - If k is too small, the approach is sensitive to noise
 - If k is too large, the neighborhood can include examples belonging to other classes

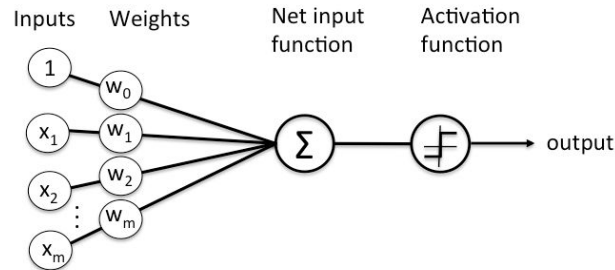
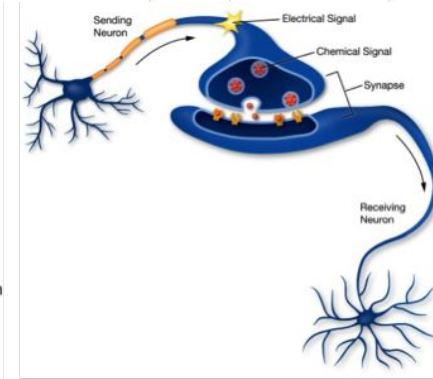
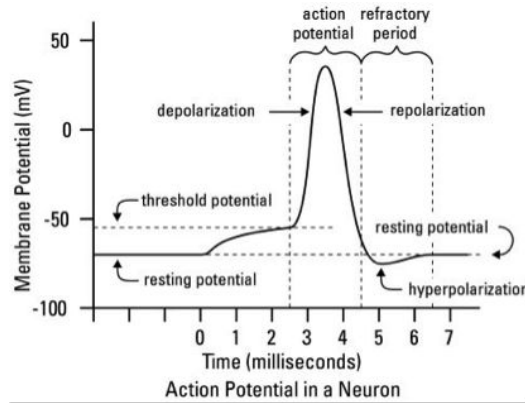


NEURAL NETWORKS

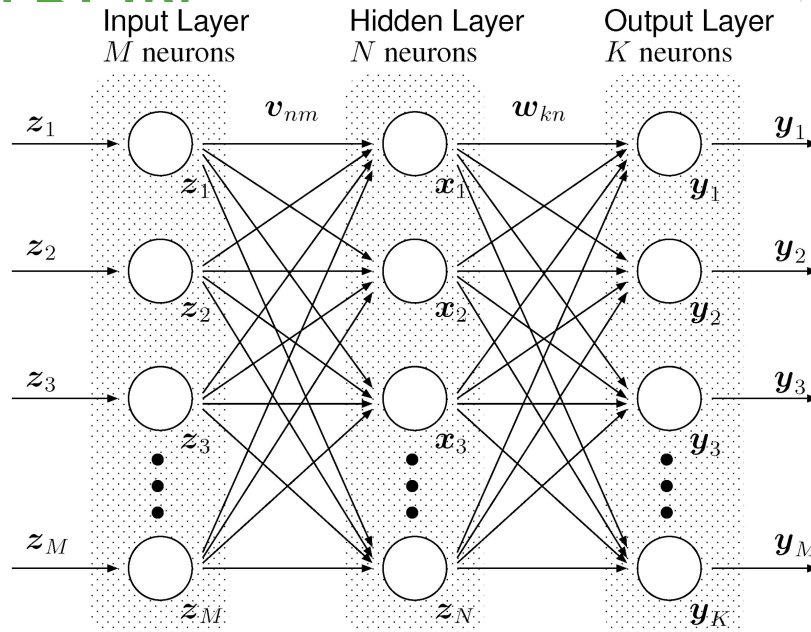
Artificial neural networks are algorithms that learn from examples using an attempt to reproduce the behavior of the human brain. These networks have historically experienced three waves:
Cybernetics (1940-1960): study of single biological neurons
Connectionism (1980-1990): study of neural networks consisting of multiple neurons
Deep Learning (2000 - present): use of neural networks with many levels of neurons




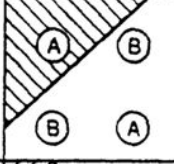
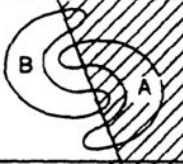

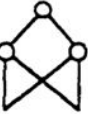
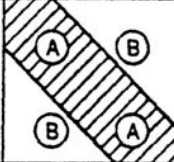
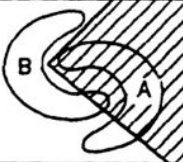


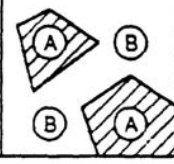
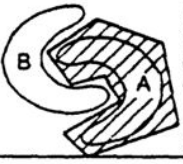

CIBERNETICA - ROSENBLATT'S PERCEPTRON



CONNESSIONISMO - MULTILAYER PERCEPTION

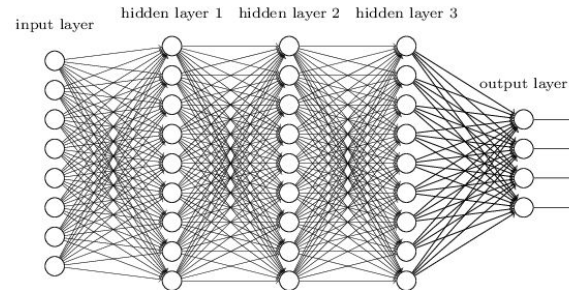
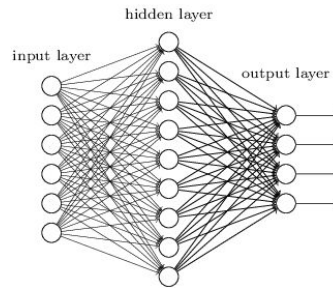


LE REGIONI DI DECISIONE DI UNA RETE NEURALE

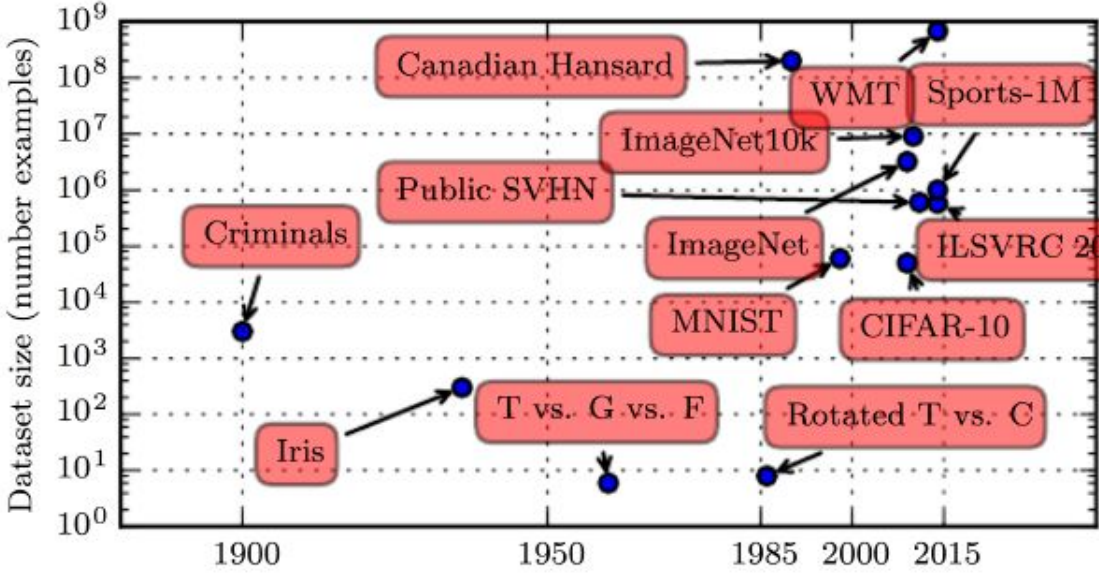
Structure	Type of Decision Regions	Exclusive-OR Problem	Classes with Meshed Regions	Most General Region Shapes
Single-layer 	Half plane bounded by hyperplane			
Two-layers 	Convex open or closed regions			
Three-layers 	Arbitrary (Complexity limited by number of nodes)			

LIMITATIONS OF CLASSICAL NEURAL NETWORKS

The defects of the neural networks used in the 90s are related to the reduced power of the computers of the time and to small datasets. Networks with one or two hidden layers and low total number of neurons: inability to solve very complex problems. With the increase in computing power (and GPUs) and with large datasets, the very popular deep learning was born.

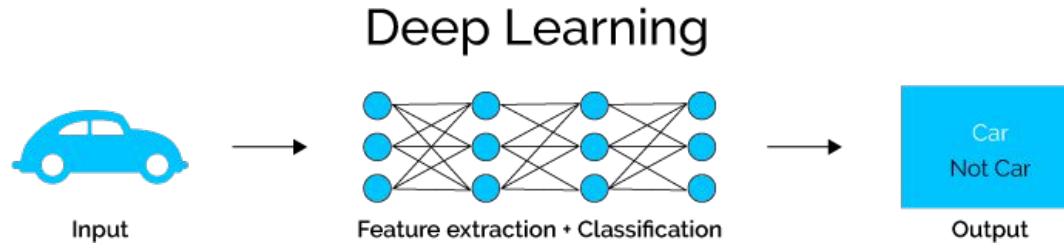
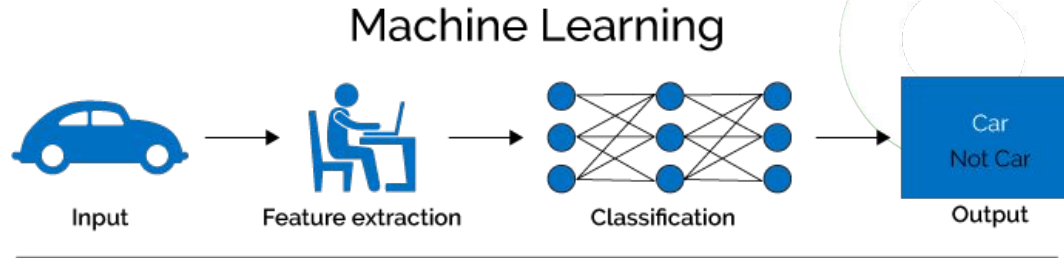


DEEP DATASET



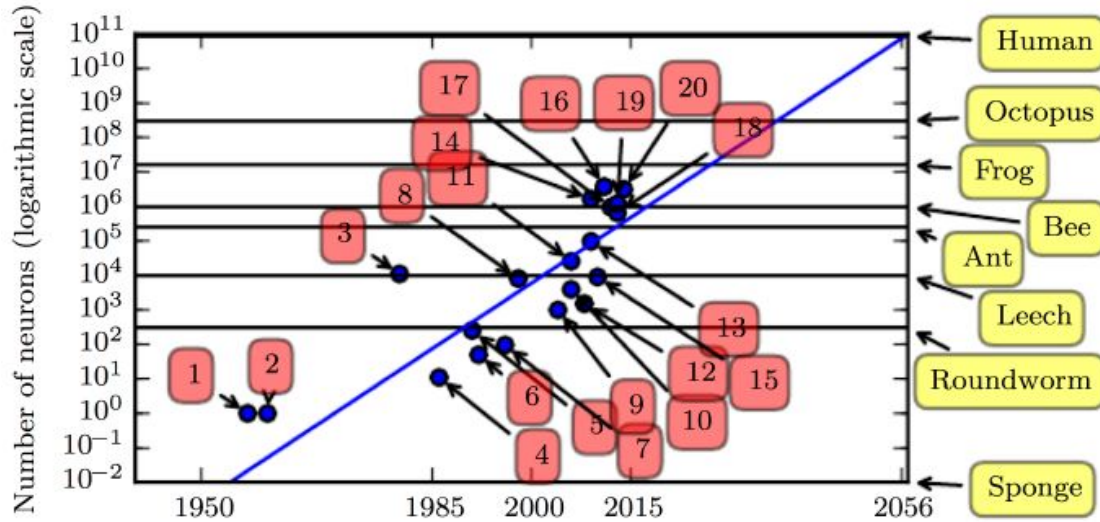
DEEP LEARNING

With deep learning, features are learned "automatically" from the data, with a progressive automatic process that builds an increasingly adequate representation of the objects of interest.



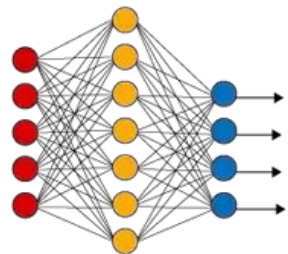
DEEP NETWORKS

A projection on the growth of the number of neurons in artificial neural networks proposed in the literature would suggest that in 2056 these could have the same number of neurons as the human brain.

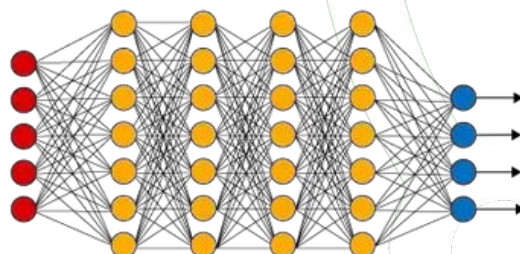


CONVOLUTIONAL NEURAL NETWORKS (CNN)

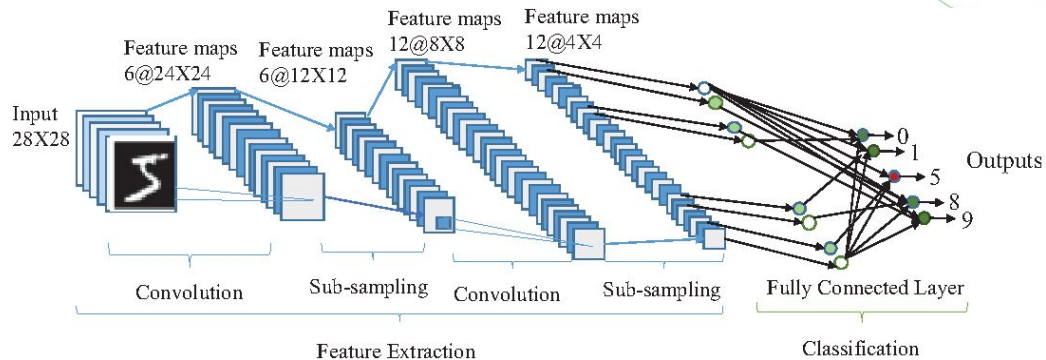
Simple Neural Network



Deep Learning Neural Network



● Input Layer ● Hidden Layer ● Output Layer



PRESTAZIONI SUL DATASET IMAGE NET

In recent years, some CNNs such as AlexNet, VGG, GoogLeNet and ResNet have achieved comparable or even superior performance compared to humans on a problem of classifying images into 1000 categories.

