UNIVERSITÀ DEGLI STUDI DI MILANO

Artificial Intelligence

History of AI

The field of AI research was born at a workshop at Dartmouth College in 1956 where the term "Artificial Intelligence" was coined by John McCarthy to avoid restricting the term to fields such as cybernetics.

At the workshop, researchers from several disciplines met to clarify, define ideas and establish a research program concerning “**thinking machines**”.

They chose the name “Artificial Intelligence” for its broad sense, to avoid restricting the interests of this field to subjects such as cybernetics, automata theory and complex information processing.

“Dartmouth Summer Research Project on Artificial Intelligence” is now considered by many [1], [2] the seminal event where Artificial Intelligence (AI) was officially declared a research field.

[1] R. J. Solomonoff, "Artificial intelligence social effects future developments", *Hum. Syst. Manage.*, vol. 32, pp. 149-153, 1985.

[2] J. Moore, "The dartmouth college artificial intelligence conference: The next fifty years", *AI Mag.*, vol. 27, no. 4, pp. 87-91, 2006.

DEFINITIONS

Today, AI:

(Merriam-Webster Dictionary:) “concerns the theory and development of computerized systems able to imitate and simulate human intelligence and behavior”

essentially “being human-like rather than becoming human” [3], and

(English Oxford Living Dictionary:) “performing tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages”

[3] B. Marr, The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance, Feb. 2018, [online] Available: https://www.tobepublishedforbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance.

Computer science studies defitions

[4,5,6,7] Define AI researches (called “computational intelligence” researches by some authors [4]):

studies about “intelligent agents” or “rational agents [5]”: any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.

[4] Poole D. and Mackworth A. and Goebel, R. (1998). “Computational Intelligence: A Logical Approach”. New York: Oxford University Press. ISBN 978-0-19-510270-3.

[5] Russell, Stuart J.; Norvig, Peter (2003), Artificial Intelligence: A Modern Approach (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2.

They use the term rational agents instead of intelligent systems and write (page. 55): "The whole-agent view is now widely accepted in the field"

[6] Jackson, Philip (1985): Introduction to Artificial Intelligence (2nd ed.). Dover. ISBN 978-0-486-24864-6.

[7] Legg S. and Hutter M. (15 June 2007). A Collection of Definitions of Intelligence (Technical report). IDSIA. arXiv:0706.3639. Bibcode:2007arXiv0706.3639L. 07-07.

In [8] AI is defined in a way that seems introducing the field of machine learning:

a (computational agent) “system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation”.

[8] Kaplan A. and Haenlein M. (2019), "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence". *Business Horizons*. **62** (1): 15–25. doi:10.1016/j.bushor.2018.08.004

UNIVERSITÀ DEGLI STUDI DI MILANO

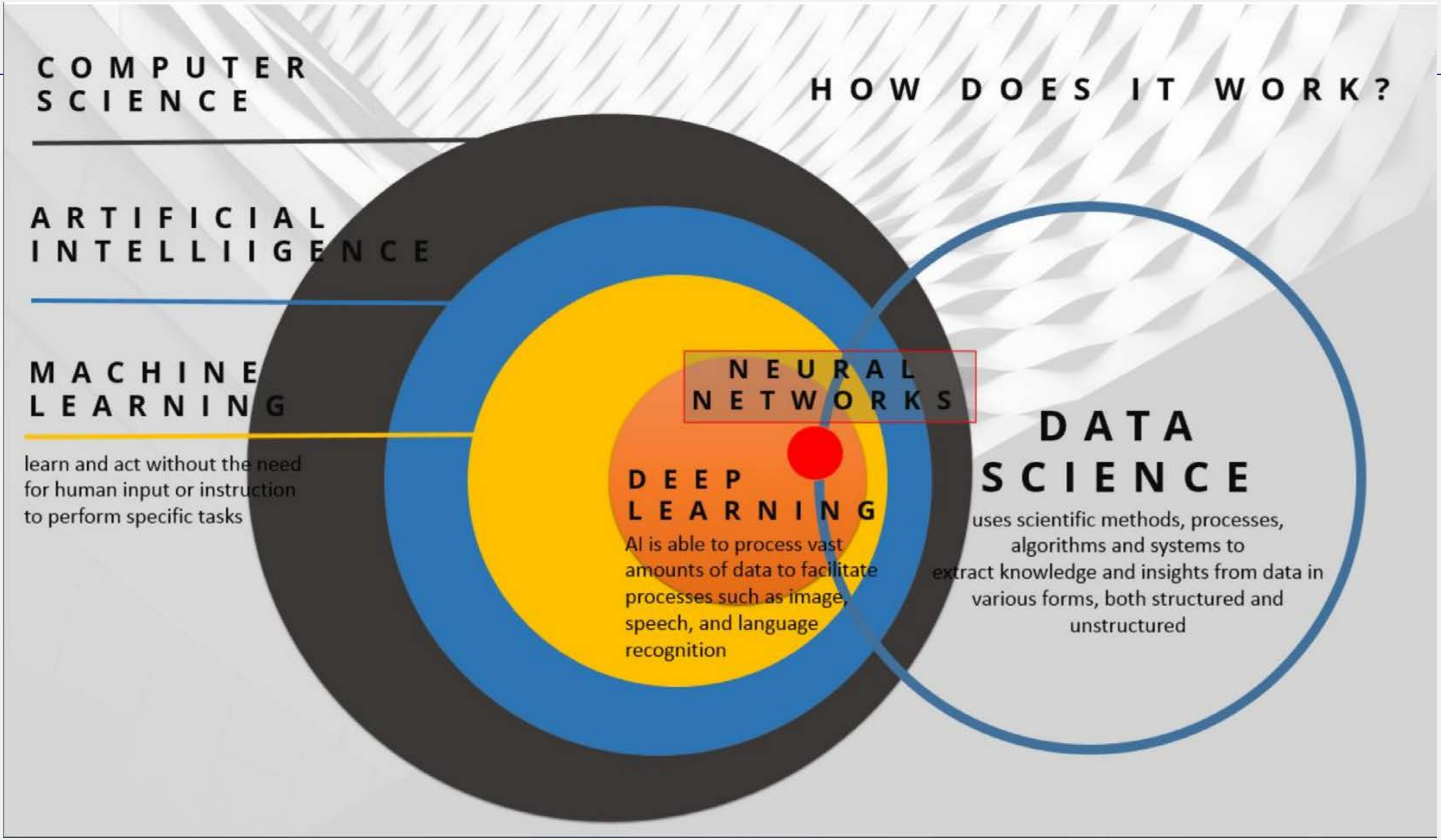
“Artificial intelligence (AI) can be described as the ability of a computer or robot-controlled computer to perform tasks that are commonly associated with intelligent creatures”

“scientific discipline that involves building computer systems whose behavior can be interpreted intelligently”

Catching-up

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence implemented into, and demonstrated by, machines

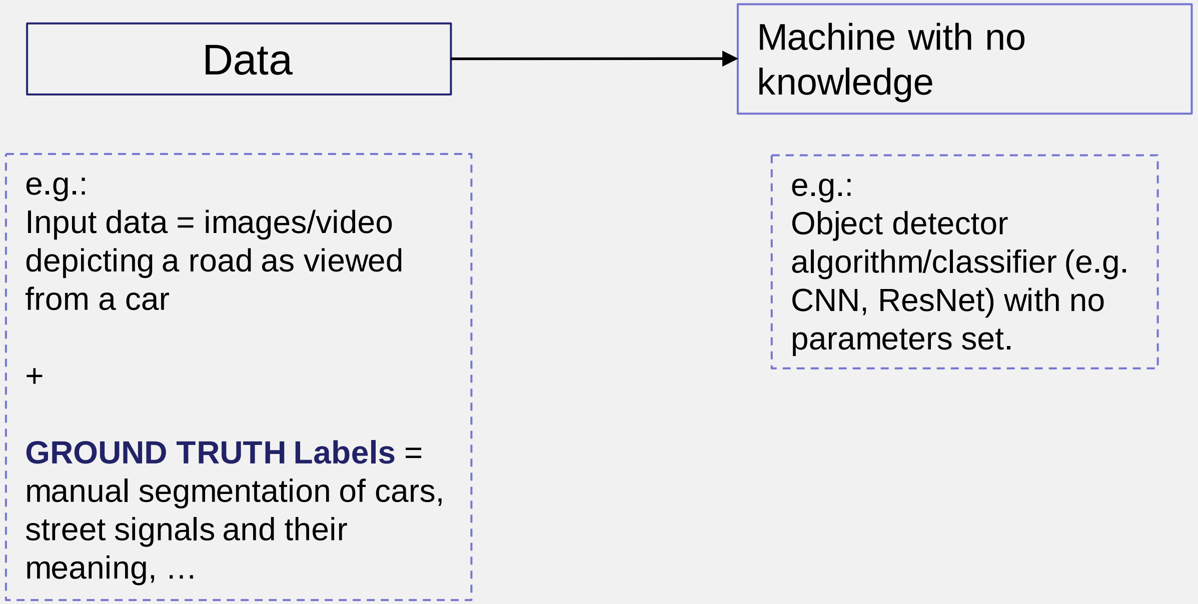
It differs from Natural intelligence, which is the one displayed by humans and other animals


COMPUTER SCIENCE
HOW DOES IT WORK?
ARTIFICIAL INTELLIIGENCE
MACHINE LEARNING
learn and act without the need for human input or instruction to perform specific tasks
NEURAL NETWORKS
DEEP LEARNING
AI is able to process vast amounts of data to facilitate processes such as image, speech, and language recognition
DATA SCIENCE
uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured

Machine Learning (Supervised learning)

AI is formed through learning (by viewing examples).

Once the machine has learned it can view novel (unknown – never seen) data to generate its own opinions (in the form of predictions or classifications).



Data is split into Training and Test sets

Training set:

Training Set

Validation Set

→ Machine training: learning algorithms choose the best machine setting

Learning refers to choosing the algorithm parameters that allow achieving the best performance on both the training and the validation set

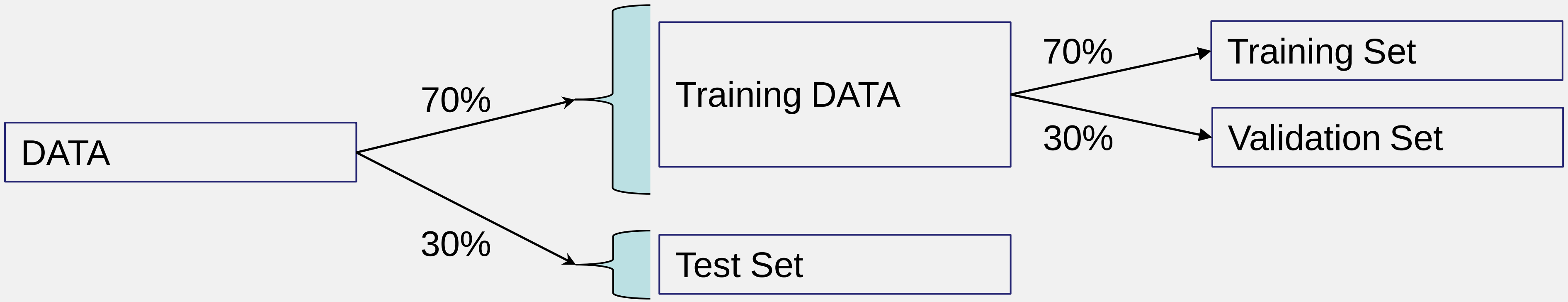
Test Set:

Test Set → AI (trained) Machine Testing

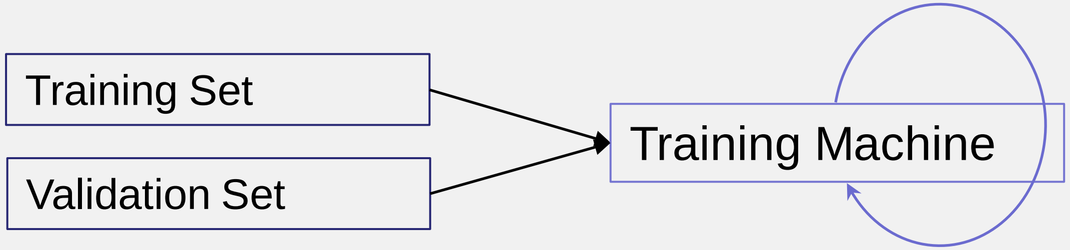
Testing refers to the application of the trained machine to unseen data contained in the test set.

Training data is further slit into:

- Training set (for training the machine through iterative methods)
- Validation set (for validating the machine and interrupt learning)



(Supervised) Training: a simplified training algorithm



FOR Epoch = 1:N — *The algorithm executes N iterations (epochs); generally N = 1000*

- o For each subsets of the training set: — *In each epoch, all the subsets composing the training set are analyzed.*
    - compute predictions for all the points in the subset — *Predictions are computed by using the machine at its current status*
    - measure the prediction error — *Remember the training set contains the ground truth labels*
    - Adapts the algorithm parameters to diminish the error — *Different techniques for doing so..*

  END
- o compute prediction for all the examples in the validation set and evaluate the error in the validation set.
- o IF the error on the validation set grows or hasn’t been changing for a while TRAINING STOPS;

  ELSE training continues with the next epoch

END

(Supervised) Training

We call it supervised learning because our example have GROUND TRUTH labels

Supervised learning is applied to:

- Neural models (multilayer perceptrons - MLPs, feed forward neural networks – FFNNs)
- Deep neural network
- Support Vector Machines (SVMs)
- Bayesian Trees (BTs) and Random Forests (RFs)
- ….

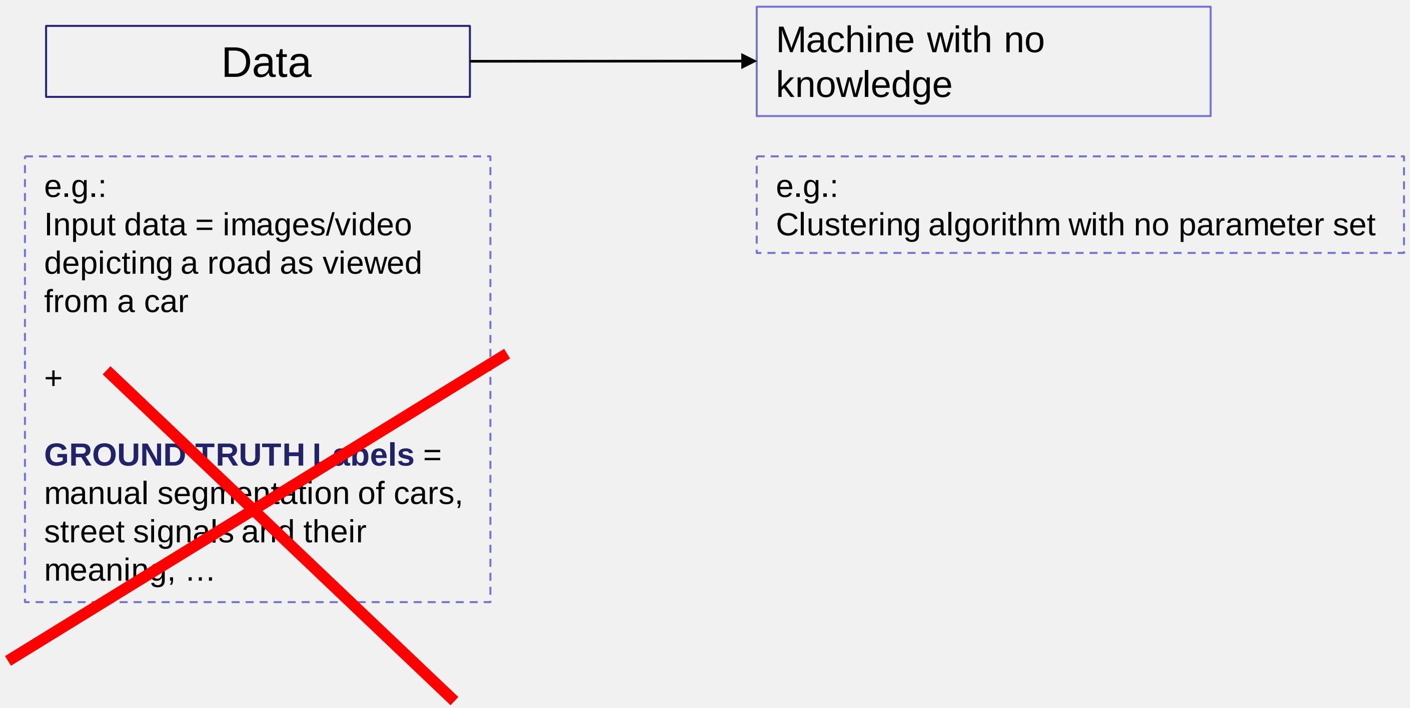
And many other (combinations) of them

https://github.com/LucaCappelletti94/bioinformatics_practice

Machine Learning (UNSUPERVISED learning)

AI is formed through learning (by viewing examples WITHOUT GROUND TRUTH LABELS).

Once the machine has learned it can view novel (unknown – never seen) data to generate its own opinions (in the form of predictions or classifications).



AGAIN

The dataset is split

The algorithm parameters are estimated on the training set

The algorithm is tested on the test set

Again Data is split into Training and Test sets

Training set:

Training Set

Validation Set

Machine training: learning algorithms choose the best machine setting

Learning refers to choosing the algorithm parameters that allow achieving the best performance on both the training and the validation set

Test Set:

Test Set

AI (trained) Machine Testing

Testing refers to the application of the trained machine to unseen data contained in the test set.

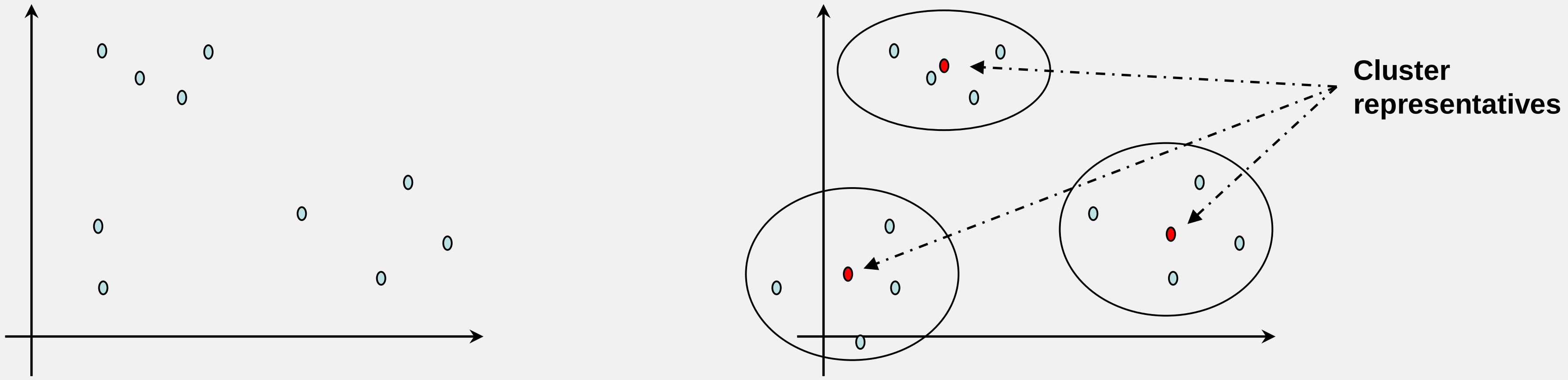
Here training requires setting the parameters of, e.g clustering algorithms.

Clustering algorithms analyze the data to form groups.

Examples, Arbib, k-means.

Example of parameters required:

- number of clusters to search for
- cluster initialization
- learning rate



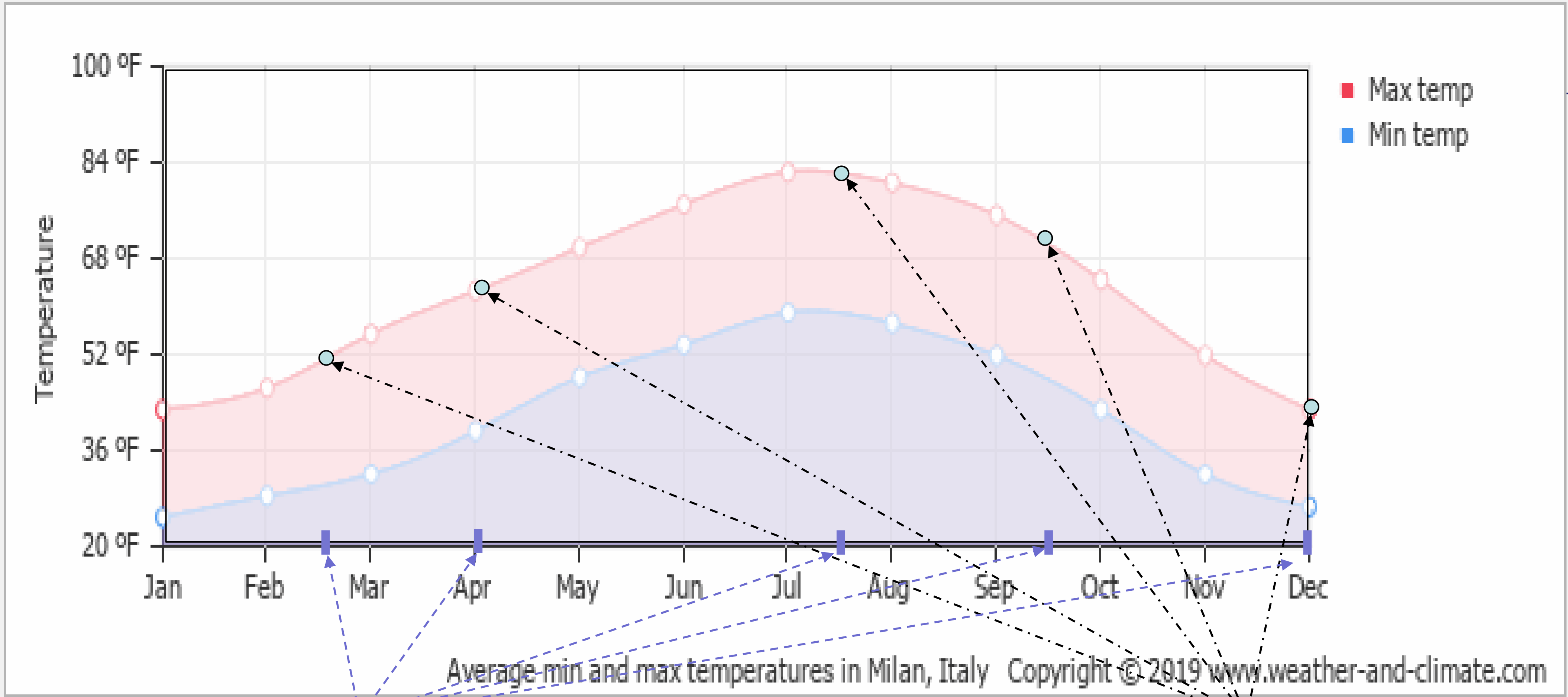
When the machine learns how to predict classes for each input data we call it Classifier.

Learning a Classification “means” somehow learning a data grouping.

When the machine must infer a continuous number from the data, we call it regressor (regression algortihms are used for training).

Learning a Regression model means learning a function which associate to each input point e function.

e.g.: given date, learn the expected temperature in Lombardy (Italy)



Points:
17th February
2nd of April
15th of July
13th of September
31st of December

Labels (for each point are the correct temperature estimates for each day)
52° F
60° F
80° F
70° F
45° F

Today we experience with a genomic classification problem

https://github.com/LucaCappelletti94/bioinformatics_practice

Points: genome sequence

4 classes: Active Enhancers, Active Promoters, Inactive Enhancers, Inactive Promoters

DNA

Genes influence what we look like on the outside and how we work on the inside.
They contain the information our bodies need to make chemicals called proteins.
Proteins form the structure of our bodies, as well playing an important role in the processes that keep us alive.

Genes are made of a molecule (chemical) called **DNA**, which is short for 'deoxyribonucleic acid'.
The **DNA molecule is a double helix**: that is, two long, thin strands twisted around each other like a spiral staircase.

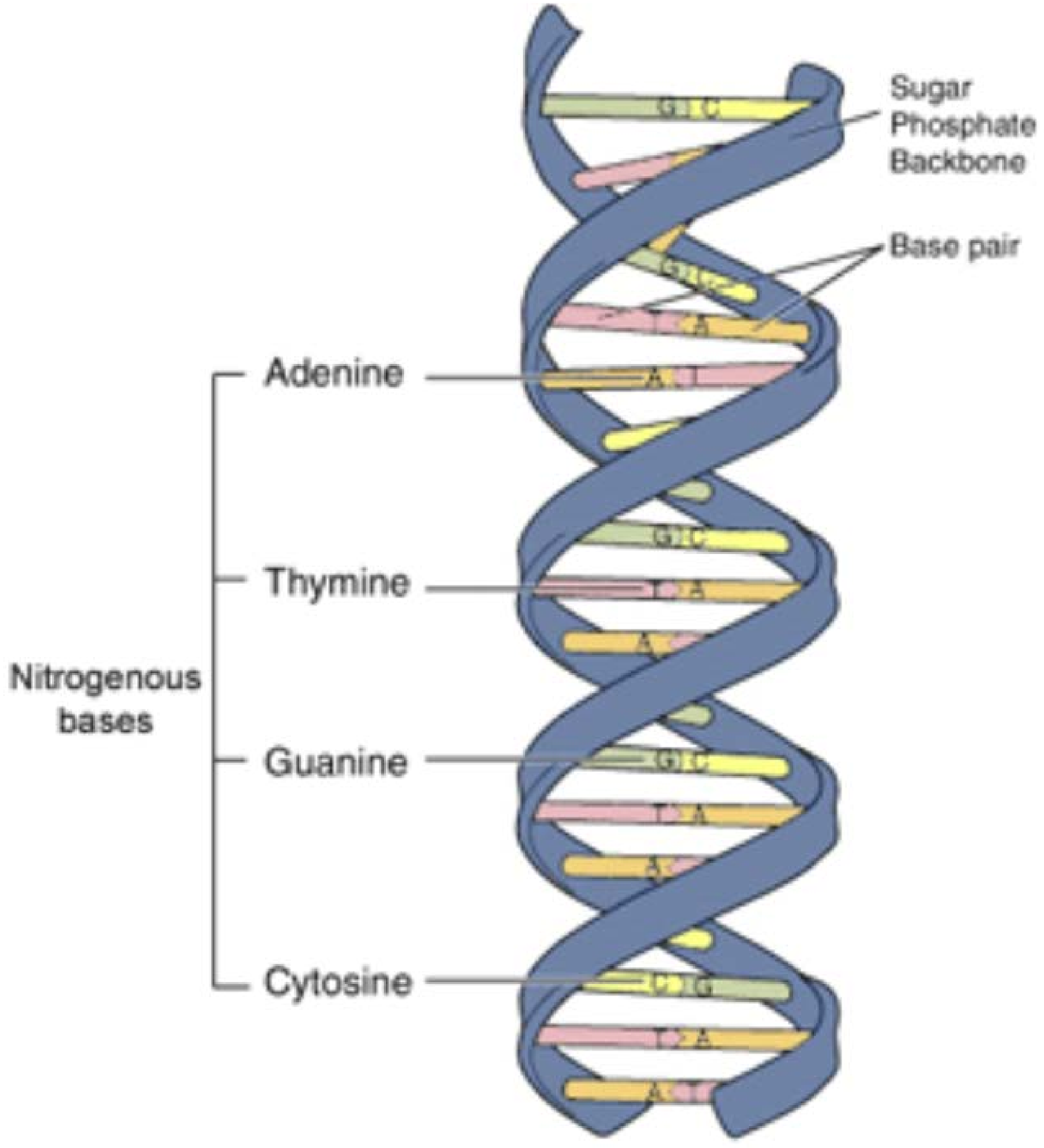


Image adapted from: National Human Genome Research Institute.

The DNA double helix showing base pairs

The sides are sugar and phosphate molecules.

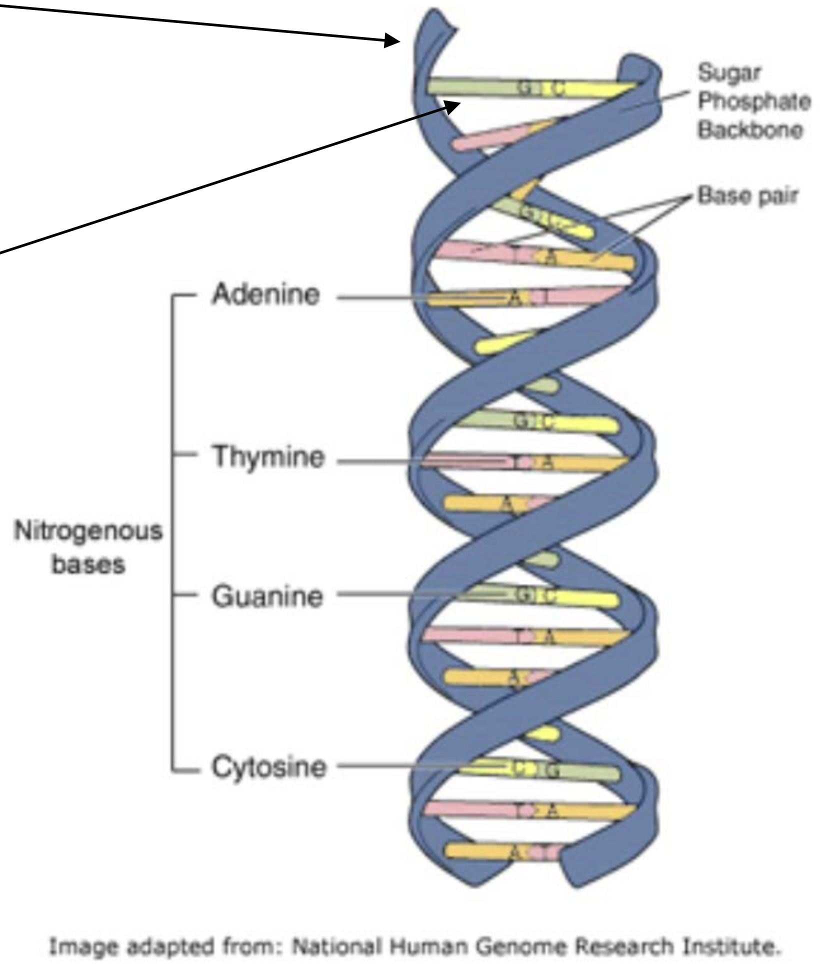
The rungs are pairs of chemicals called **'nitrogenous bases'**, or **'bases'** for short.

There are four types of base: adenine (A), thymine (T), guanine (G) and cytosine (C).

These bases link in a very specific way: A always pairs with T, and C always pairs with G.

The DNA molecule has two important properties.

- **It can make copies of itself**. If you pull the two strands apart, each can be used to make the other one (and a new DNA molecule).
- **It can carry information**. The order of the bases along a strand is a code - a code for making proteins.

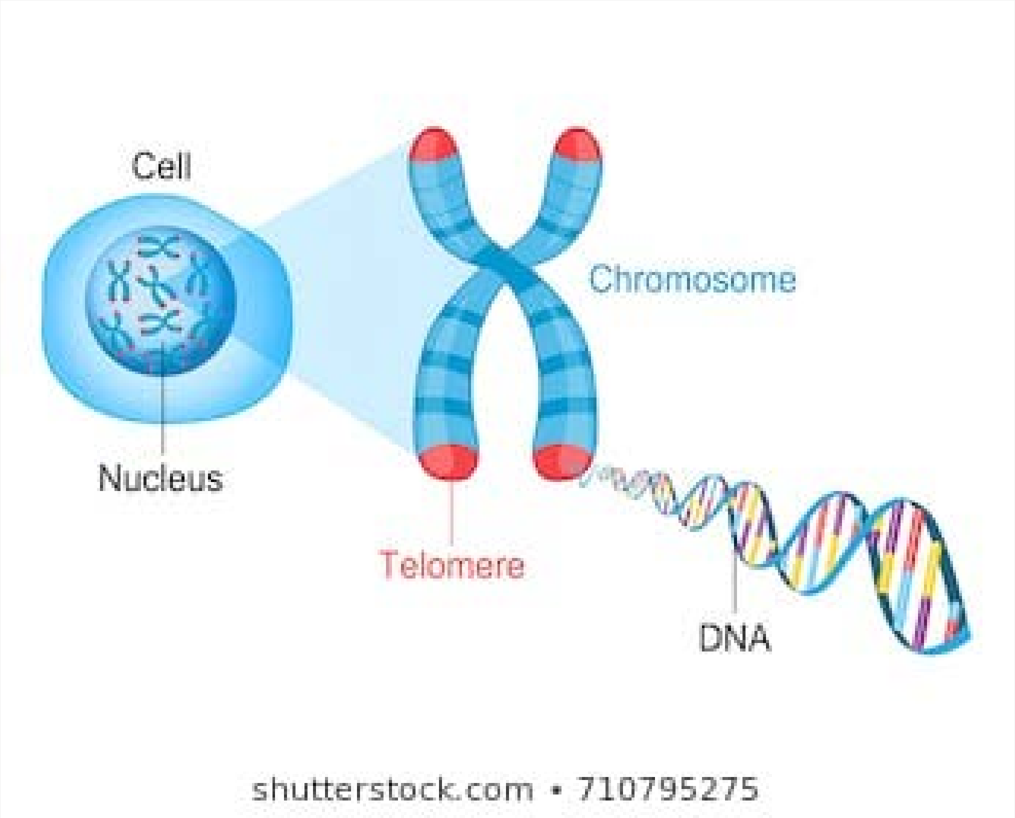


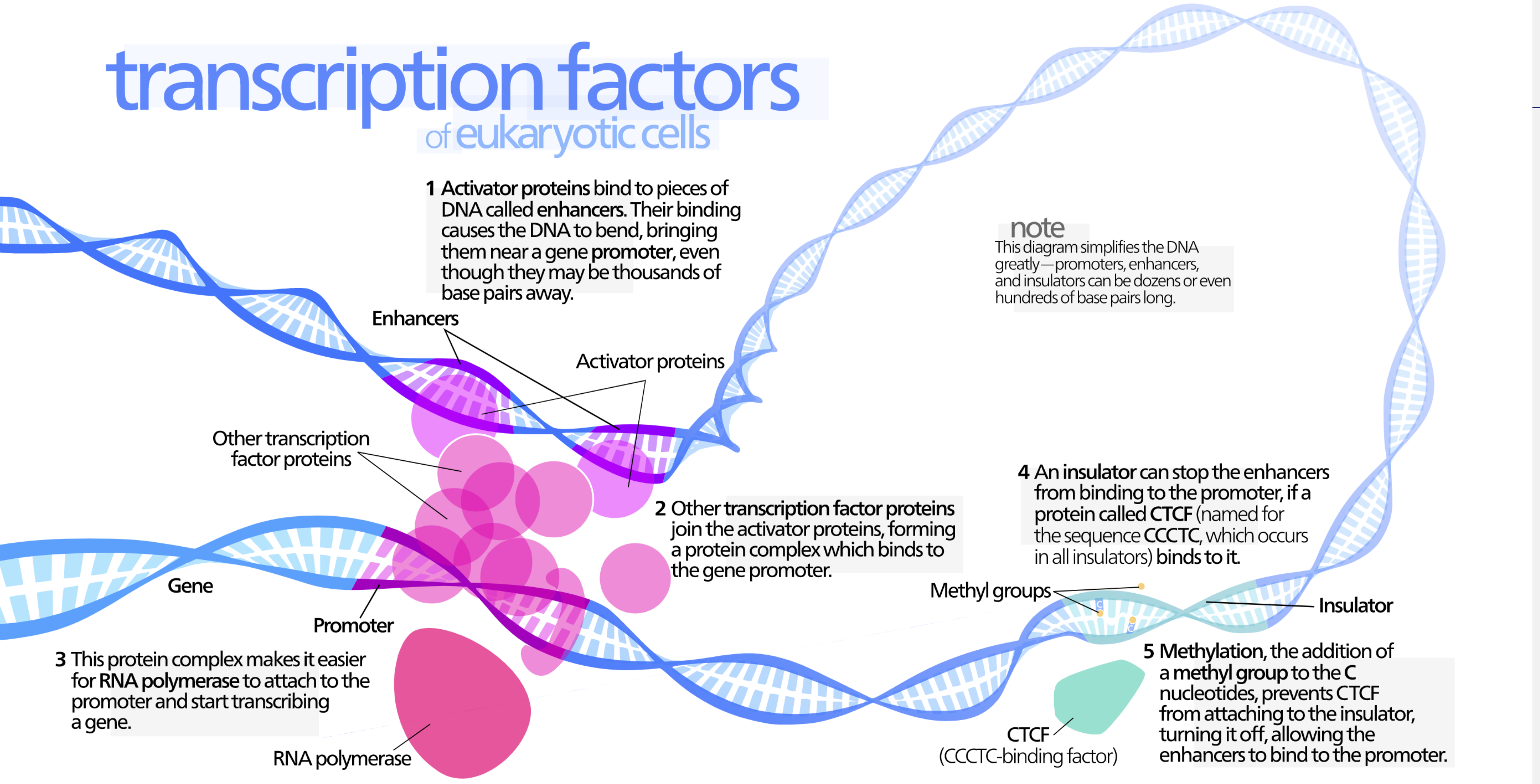
The DNA double helix showing base pairs

UNIVERSITÀ DEGLI STUDI DI MILANO

A **gene** is a length of DNA that codes for a specific protein. So, for example, one gene will code for the protein insulin, which is important role in helping your body to control the amount of sugar in your blood.

Chromosomes are 46 (types) in each individual, they are long filaments of genes and proteins with an x-shape.




transcription factors
of eukaryotic cells
1 Activator proteins bind to pieces of DNA called enhancers. Their binding causes the DNA to bend, bringing them near a gene promoter, even though they may be thousands of base pairs away.
note
This diagram simplifies the DNA greatly—promoters, enhancers, and insulators can be dozens or even hundreds of base pairs long.
Enhancers
Activator proteins
Other transcription factor proteins
2 Other transcription factor proteins join the activator proteins, forming a protein complex which binds to the gene promoter.
4 An insulator can stop the enhancers from binding to the promoter, if a protein called CTCF (named for the sequence CCCTC, which occurs in all insulators) binds to it.
Gene
Promoter
Methyl groups
Insulator
3 This protein complex makes it easier for RNA polymerase to attach to the promoter and start transcribing a gene.
RNA polymerase
CTCF
(CCCTC-binding factor)
5 Methylation, the addition of a methyl group to the C nucleotides, prevents CTCF from attaching to the insulator, turning it off, allowing the enhancers to bind to the promoter.

We describe each genetic sequence (POINT) of active/inactive enhancer/promoter through **epigenomic data**

??? EPIGENOMIC DATA ???

Genotype vs Phenotype

Before, we must see what genotype and phenotypes are

An organism's genotype is the set of genes that it carries.

An organism's phenotype is all of its observable characteristics — which are influenced both by its genotype and by the environment.

So in defining evolution, we are really concerned with changes in the genotypes that make up a population from generation to generation.

However, since an organism's genotype generally affects its phenotype, the phenotypes that make up the population are also likely to change.

For example, differences in the genotypes can produce different phenotypes. In these house cats, the genes for ear form are different, causing one of these cats to have normal ears and the other to have curled ears.

A change in the environment also can affect the phenotype. Although we often think of flamingos as being pink, pinkness is not encoded into their genotype. The food they eat makes their phenotype white or pink

UNIVERSITÀ DEGLI STUDI DI MILANO

Epigenomics

Epigenomics is the study of the complete set of epigenetic modifications on the genetic material of a cell (epigenome).

Epigenetic modifications are reversible modifications on a cell's DNA or histones that affect gene expression without altering the DNA sequence.

For each sequence, epigenomic data express how much that sequence is involved is a list of genes-protein interactions.