Machine Learning

Introduction

Parcours Progis

Etudes, Medias, communication, Marketing

11.09.2025

Bahareh Afshinpour

UGA Université Grenoble Alpes

Sciences Po Grenoble UGA

Course Information

- **Bahareh Afshinpour**
- **Email:** bahareh.afshinpour@univ-grenoble-alpes.fr
- **Home page:** http://afshinpour.com/
- **Office:** IMAG building -371
- **Course time**
  - 1h30 TD
  - 1h30 TP

Textbook and References

- **Ethem Alpaydin, ”Introduction to Machine Learning”, MIT Press, Prentice Hall of India, 3rd 2014.**
- **Neural Networks for Pattern Recognition by C. Bishop**
- **Python Machine Learning by S. Raschka and V. Mirjalili (highly recommended).**
- **YouTube videos.**

Skills you will acquire

After completing this course, you will be able to:

- ✓ explain, compare, and contrast various machine learning topics and concepts like supervised learning, unsupervised learning, classification, regression, and clustering.

- ✓ You will also be able to describe how the various machine learning algorithms work.

- ✓ And finally, you will learn how to apply these machine learning algorithms in Python using various Python libraries.

Grading

- TP (Homeworks- Report) 25%
- Mid-Term Exams 25%
- Final Exam 50%
- (Final project)

Policies

- Attendance is required.
- All submitted work must be yours.

- Protect your efforts! Don't let others see your codes, don't give others your results.
  - Lending your codes to others or allowing others to copy your work will be considered as collusion, thus receiving the same punishment as the plagiarist.

- Your codes need to be able to generate the results.

Computing needs

- **Basic Desktop or Laptop**

- **Ability to :**
  - **Install and run Python + notepad++**
    - » (Python offers a large number of packages that make it simple to get started and build applications of all complexity levels.)
  - **Jupyter Notebook**
    - » **(**The Jupyter Notebook is a simple interactive environment for programming with Python, which makes it really easy to share your results.**)**

Introduction

- **This is the age of "big data"**
  - **we all became producers of data.**
    - ✓ Every time we buy a product, write a blog, or post on the social media, we are generating data.

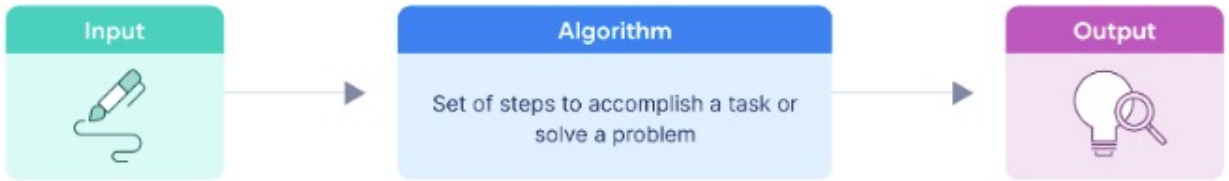
https://www.lebigdata.fr/definition-big-data

  - **Each of us is not only a generator but also a consumer of data.**

✓We want to have products and services specialized for us.
✓We want our needs to be understood and interests to be **predicted**.
✓There are certain **patterns** in the data.

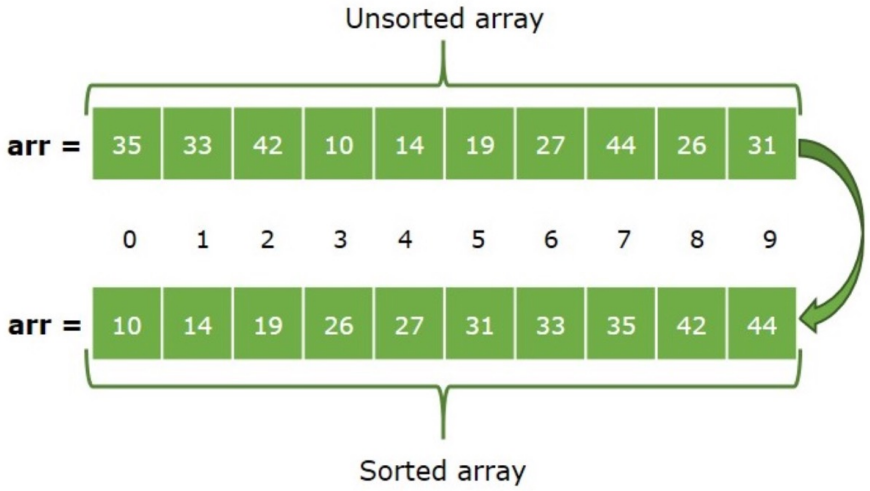
Introduction- Algorithm

- **To solve a problem on a computer, we need an algorithm.**



https://www.scribbr.com/ai-tools/what-is-an-algorithm/

- **An algorithm is a sequence of instructions that should be carried out to transform the input to output.**
  - ✓ **For example, one can devise an algorithm for sorting. The input is a set of numbers and the output is their ordered list.**



https://www.tutorialspoint.com/python/python_sort_arrays.htm

Introduction- Algorithm

- **For some tasks, however, we do not have an algorithm.**
  - ✓ **Predicting customer behavior is one;**
  - ✓ **another is to tell spam emails from legitimate ones.**
    - **what we want is to "learn" what constitutes spam from them.**

In other words, we would like the computer(machine) to extract automatically the algorithm for this task.

Any piece of software that will consume training examples,

in order to make decisions over ***unseen*** data

without explicit programming is considered

learning.

Introduction- Machine Learning application

- **Machine Learning application areas are abundant:**
  - **In finance banks**
    - Fraud detection
    - Credit approval
    - Price and market prediction
  - **In manufacturing,**
    - learning models are used for optimization, control, and troubleshooting.
  - **In medicine,**
    - drug discovery
    - computational genomics (analysis and design)
    - medical imaging and diagnosis
  - **In telecommunications,**
    - call patterns are analyzed for network optimization and maximizing the quality of service.
  - **Computer vision and robotics:**
    - detection, recognition and categorization of objects
    - face recognition
    - tracking objects (rigid and articulated) in video
  - **In science,**
    - large amounts of data in physics, astronomy, and biology can only be analyzed fast enough by computers.

Introduction- Machine Learning application

- **social media domain :**

  Example: Facebook, Instagram, TikTok, and YouTube use ML to personalize users' feeds. Algorithms analyze your **past interactions** (likes, shares, watch time) **to predict and recommend** content you're likely to engage with.

- **political domain:**

  Sentiment Analysis & Public Opinion Mining : ML models analyze social media posts, news articles, and surveys to gauge public sentiment toward policies, candidates, or political events. Tools like Brandwatch or Crimson Hexagon are used for real-time sentiment tracking.

Introduction- What Is Machine Learning?

- Machine learning is

  the process of teaching computers how to improve performance by using example data or information from the past.

The real question is what is learning?

Using past experiences to improve future performance.

For a machine, experiences come in the form of **data**.

Introduction- What Is Machine Learning?

- A model is defined by parameters. Learning involves running a program to optimize these parameters using training data or past experience.

- The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both.

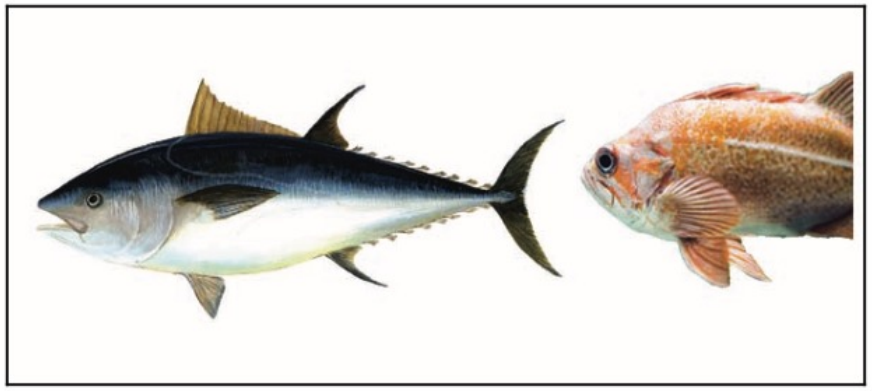
Consider a real example

Title: Automating Fish Species Classification

- **Objective:**
  - The Nature Conservancy and partners aim to monitor fishing activities and preserve fisheries using camera-based surveillance.
- **Challenge:**
  - Cameras will generate massive amounts of video data.
  - Manual review is time-consuming and expensive.

Consider a real example

- So our aim in this example is to separate different species such as tunas, sharks, and more that fishing boats catch. As an illustrative example, we can limit the problem to only two classes, tuna and opah.



Feature Extraction for Fish Classification

- **Solution:**
  - Develop a machine learning algorithm to automatically detect and classify fish species in videos.
  - Focus on scalability and efficiency for real-world deployment.
  - Binary classification: Distinguish between tuna and opah .

Key Features for Classification

Feature	Tuna	Opah
Length	Longer	Shorter
Width	Narrower	Wider
Color	Blue & white	Reddish

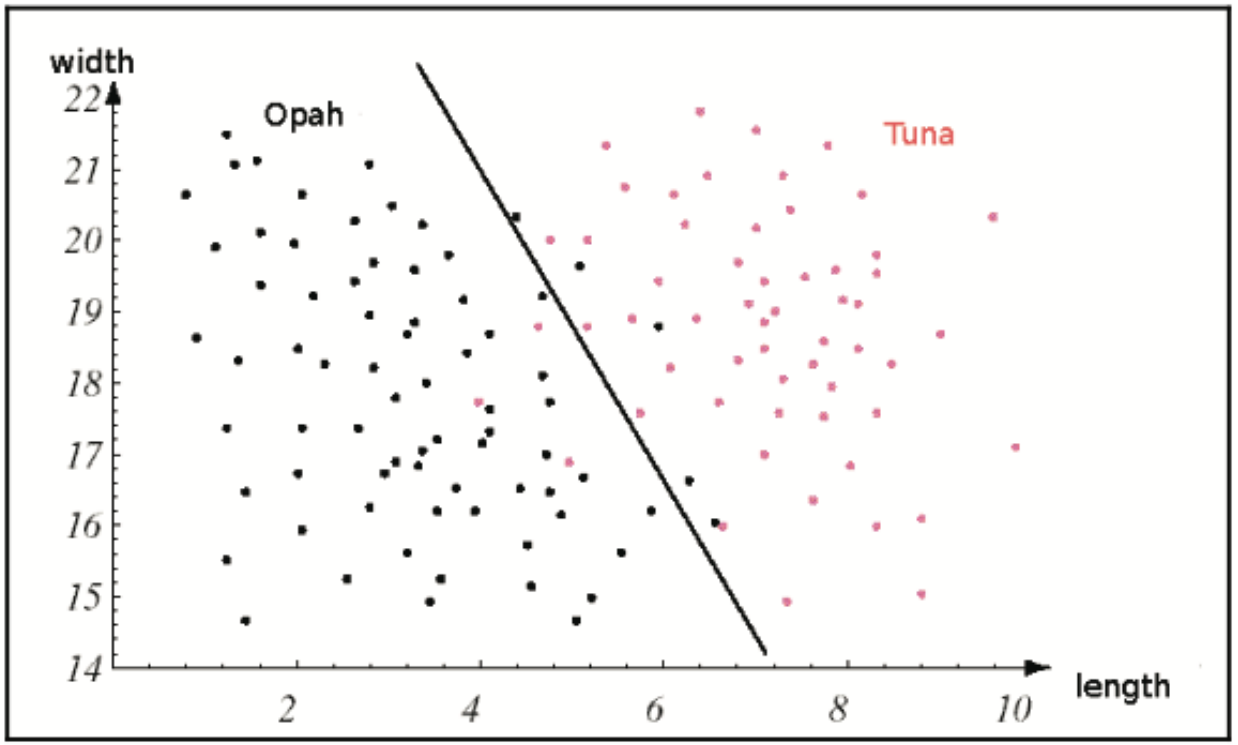
We can use these physical differences as features that can help our learning algorithm(classifier) to differentiate between these two types of fish.

Consider a real example

- In this case, the collection of tuna and opah fish will act as the **knowledge base** for our classifier.
- Initially, the knowledge base (training samples) will be **labeled/tagged**, and for each image, you will know beforehand whether it's tuna or opah fish.
- So the classifier will use these training samples to model the different types of fish, and then we can use the output of the training phase to automatically label unlabeled/untagged fish that the classifier didn't see during the training phase.
- This kind of unlabeled data is often called **unseen data**.

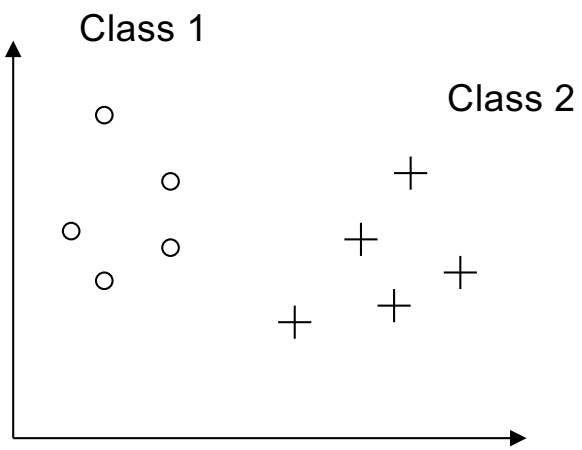
Consider a real example

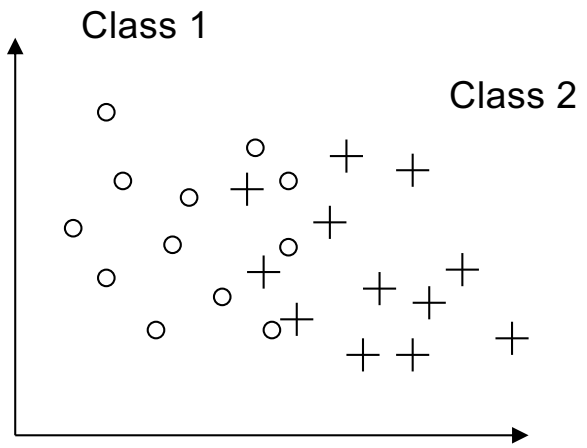
- if we combine both features, we might get something that looks like the following graph:



Feature

- Features are numerically expressed properties of the signal.
- The set of features used for pattern recognition is called feature vector.
- The number of used features is the dimensionality of the feature vector.
- n-dimensional feature vectors can be represented as points in n-dimensional feature space.



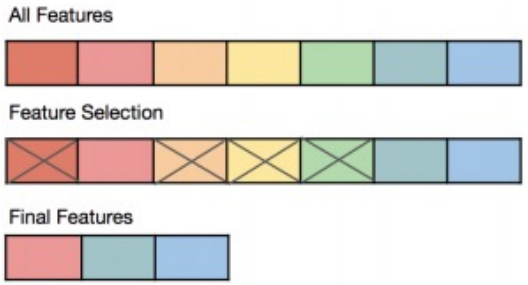


Feature selection

- For example **having macrowave** is not very help ful feature for buying house

Useful for efficiency of predictions and interpretability

Lot size
Single Family
Year built
Last sold price
Last sale price/sqft
Finished sqft
Unfinished sqft
Finished basement sqft
floors
Flooring types
Parking type
Parking amount
Cooling
Heating
Exterior materials
Roof type
Structure style



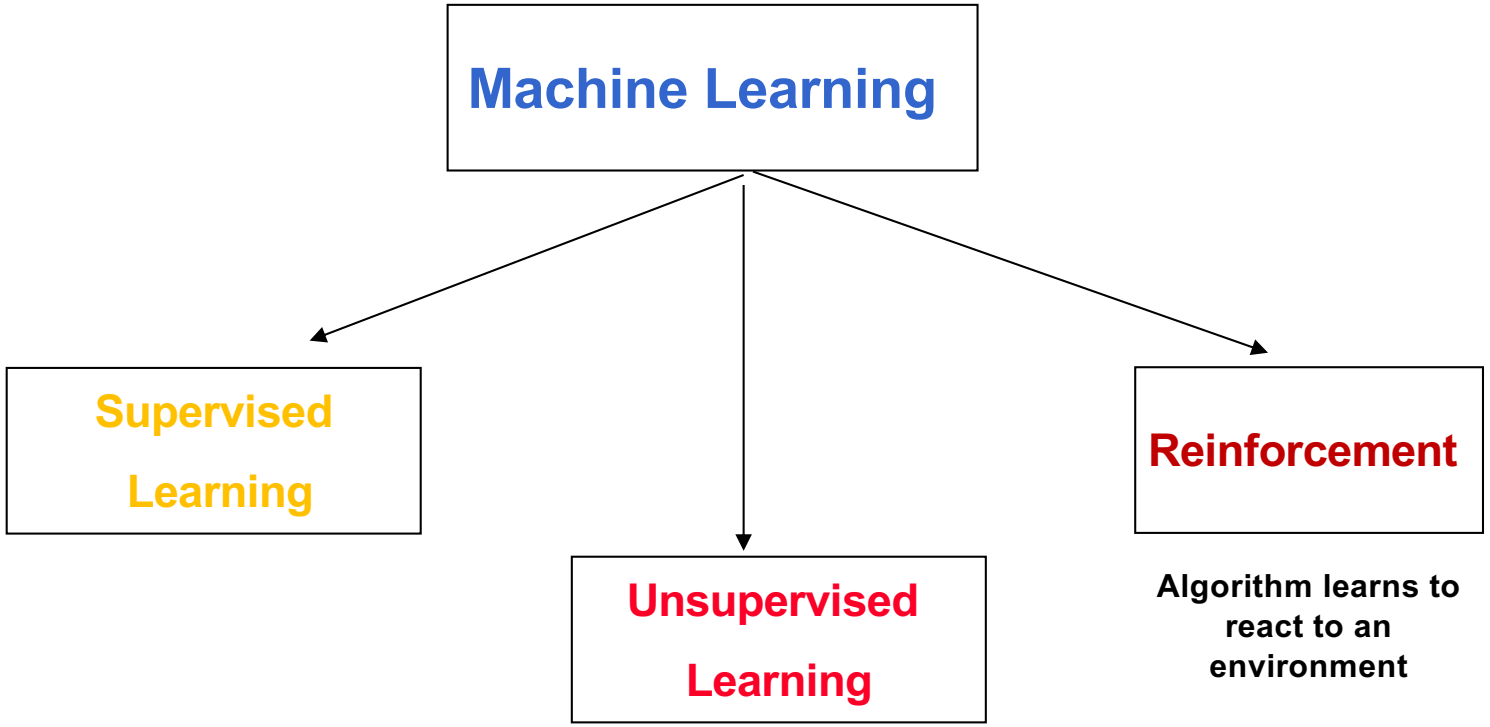
https://vitalflux.com/machine-learning-feature-selection-feature-extraction/

Feature extraction

But Feature **extraction** is about
extracting/deriving information from the original features set to create a new features subspace.

-Dimensionality Reduction

About the course



- We'll cover the most widely used machine learning methods such as
  - ✓ Regression method
  - ✓ Classification methods
  - ✓ Clustering methods
  - ✓ Random forests
  - ✓ Neural nets, and so on.

Supervised Learning Problem

Supervised Learning Problem

How do we supervise a machine learning model?

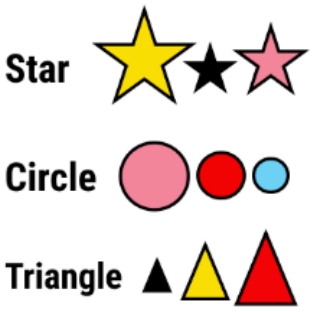
We do this by teaching the model

How exactly do we teach a model?

We teach the model by training it with some data from a labeled dataset

what does a labeled dataset look like?

It could look something like this



ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

Supervised Learning Problem

- In supervised learning,
  - we are given a data set and already know what our right output should be
  - with the assumption that there is a relationship between the input and the result
  - you have to study in relation with a different set of features called independent features.
- Two types of supervised learning tasks:
  - "Regression"
  - "classification“

Regression	Classification
Target: Continuous data	Target: Categorical data
Aim: Predict the value	Aim: Predict class/class probability

How does a supervised problem look?

X1	X2	X3	X4	X5	Y1	Y2
7	1	3	1	2	50	0
2	4	7	12	9	24	1
5	9	1	10	6	16	0
8	11	2	3	4	62	1
1	5	5	7	1	12	0

- x1,x2,x3,x4 are independent features
- Y1 and Y2 are dependent features
- Y1 is continuous
- Y2 is categorical

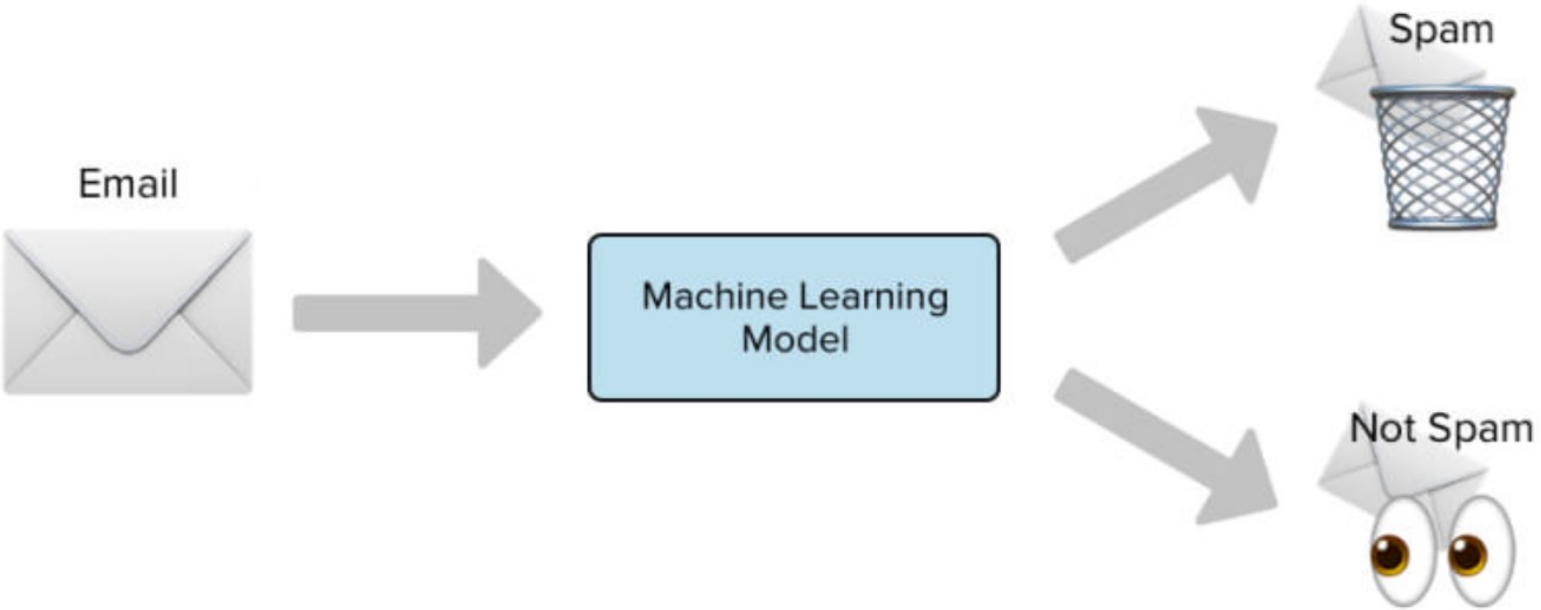
The outcome of a supervised learning problem is to have a dependent feature defined as a function of independent feature.

For example Y1=f(X1,X2,X3,X4)

How does a supervised problem look?

- **Input (Features):** Words, phrases, sender info, email length.
- **Output (Label):** "Spam" or "Not Spam".

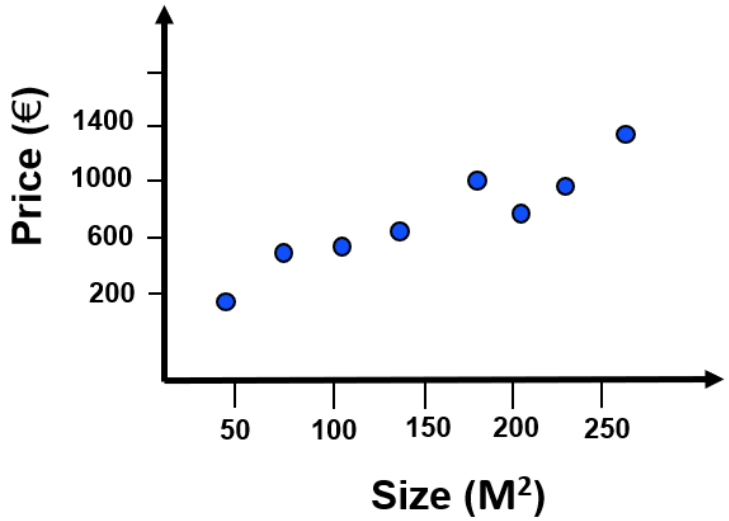
The dataset contains emails, each labeled as spam or not. The model learns patterns that distinguish spam from legitimate emails.



https://www.superannotate.com/blog/supervised-learning-and-other-machine-learning-tasks

Example of a Supervised Learning algorithm

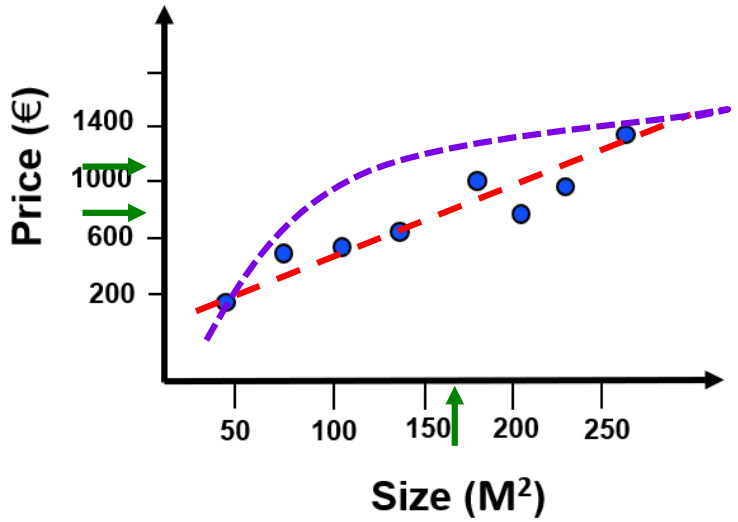
- Given data about the size of houses on the real estate market, try to predict their price.
- Price as a function of size is a **continuous** output, so this is a regression problem.



Imagine your friend has a 170 m² house and wants to know the ideal rental price. A learning algorithm can help by analyzing data from similar houses (size, location, features) and predicting a fair rental value based on past trends.

Example

- Put a straight line through the data, also fit a straight line to the data. It looks like maybe their house can be sold for about $700. But maybe this isn't the only learning algorithm you can use, and there might be a better one.



- Instead of using a straight line, we fit a curve (like a quadratic) to the data—predicting the house could rent for about $800.

What is regression?

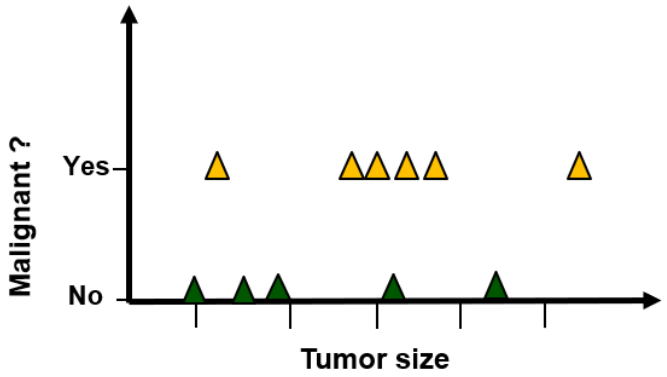
Regression is the process of predicting continuous values.

What is Classification?

Is the process of predicting discrete class labels or categories.

Examples

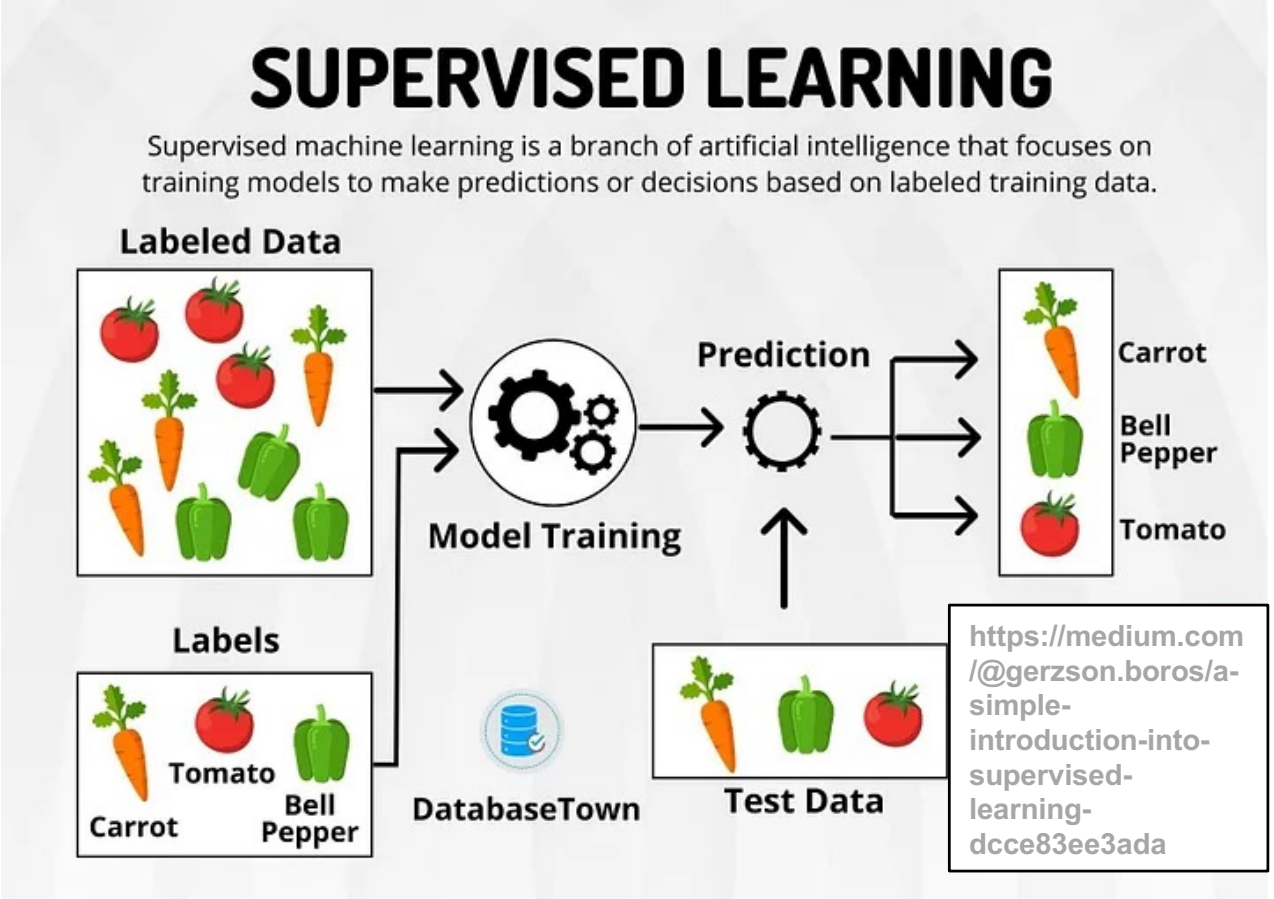
- Given a patient with a tumor, we have to predict whether the tumor is malignant or benign.
- a malignant tumor
  - is a tumor that is harmful and dangerous
- a benign tumor
  - is a tumor that is harmless



The Machine Learning question is, can you estimate what is the probability, what's the chance that tumor X as malignant versus benign?

- For some learning problems what you really want is not to use like two or three features
  - For example, the age of the patient and the size of the tumor are two features
  - instead you want to use an infinite number of features

How do you deal with an infinite number of features?


SUPERVISED LEARNING
Supervised machine learning is a branch of artificial intelligence that focuses on training models to make predictions or decisions based on labeled training data.
Labeled Data
Model Training
Prediction
Carrot
Bell Pepper
Tomato
Labels
Tomato
Carrot
Bell Pepper
DatabaseTown
Test Data
https://medium.com/@gerzson.boros/a-simple-introduction-into-supervised-learning-dcce83ee3ada

TP1

- Introduction into *Python* programming with:
  - The traditional "hello world";
  - Some data structures and variables.
  - Datatypes
  - If

WHY PYTHON

- Is a general-purpose language
- Python is one of the easier ones to learn
- There are lots of free tools out there you can use to code or learn Python
- Matured Community
- It has a large Libraries And Frameworks
- For AI, it has libraries like TensorFlow, PyTorch and Scikit-learn
- Python can also be used for Natural Language Processing using the NLTK

SO HOW DO I GET STARTED?

PYTHON LIBRARIES FOR DATA SCIENCE

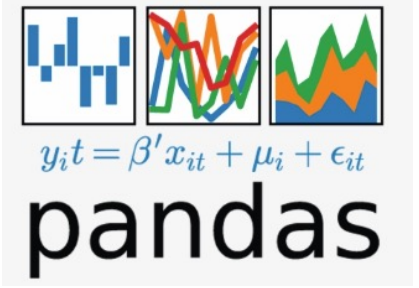
Many popular Python toolboxes/libraries:

- **NumPy**
- **SciPy**
- **Pandas**
- **SciKit-Learn**

Visualization librar

- **matplotlib**
- **Seaborn**

and many more …

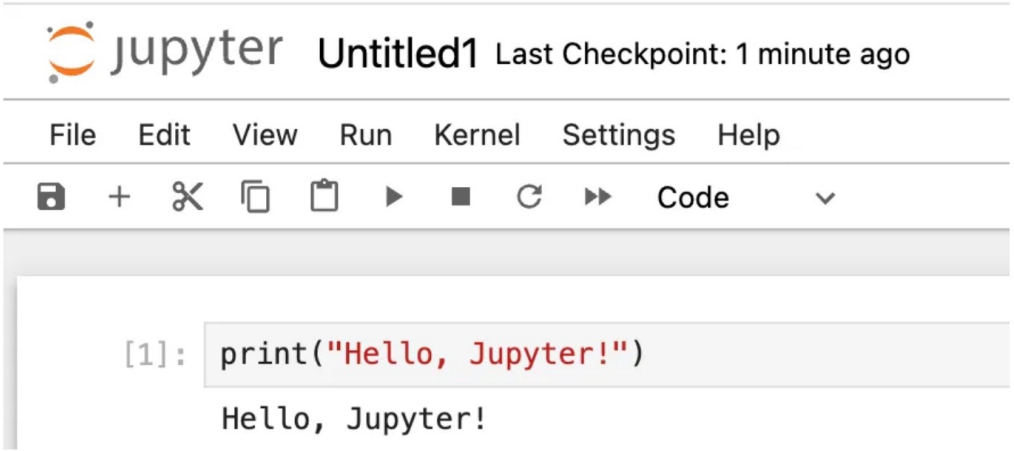


PYTHON SCIKIT-LEARN

- **Popular machine learning toolkit in Python http://scikit-learn.org/stable/**
  - **Good documentation**
  - **Is realy easy to implement**
  - **Has most of the classification, regression and clustering algorithm**
- **Requirements**
  - **Anaconda**
  - **Available from https://www.continuum.io/downloads**
  - **Includes numpy, scipy, and scikit-learn (former two are necessary for scikit-learn)**

JUPYTER

- **Jupyter is a freely available web application**
- **Jupyter promotes collaboration and reproducibility by allowing users to share their notebooks with others via email, GitHub, or the Jupyter Notebook Viewer.**



PICK UP GOOD HABITS RIGHT AWAY!

- Comments in your code help you or someone else understand
  - **What your program does**
  - **What a particular line or section of code does**
  - **Why you chose to do something a particular way**
  - **Anything that might be helpful to know if I am looking at the code later and trying to understand it!**

IN PYTHON WE USE A # TO INDICATE COMMENTS

```
#My first Python Application
#Created by me!
#Print command displays a message on the screen
print('Hello World')
```

Did you notice the colors?

ERRORS IN PYTHON

- It is important to read error messages carefully.

```
# Print string as error message

frint("Hello, Python!")
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[2], line 3
      1 # Print string as error message
----> 3 frint("Hello, Python!")

NameError: name 'frint' is not defined
```

```
# Try to see built-in error message

print("Hello, Python!)
```

```
  Cell In[3], line 3
    print("Hello, Python!)
          ^
SyntaxError: unterminated string literal (detected at line 3)
```

Python interprets your script line by line as it executes it. Python will stop executing the entire program when it encounters an error

DATA TYPES

- **A type is how Python represents different types of data.**
  - **Integers, real numbers, string and boolean**
- In Python, a string is a sequence of characters.
  - **A string can be spaces or digits. A string can also be special**
- **You can change the type of the expressio... Python, this is called typecasting.**

```
# Integer

11
```

```
# Float

2.14
```

```
# String

"Hello, Python 101!"
```

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

VARIABLES

- We can use variables to store values.
- We assign a value to a variable using the assignment operator, i.e, the equal sign.
- **We can then use the value somewhere else in the code by typing the exact name of the variable.**
- **We can use the type command in variables as well.**

PYTHON CONDITIONS AND IF STATEMENTS

Python supports the usual logical conditions from mathematics:

- **Equals: a == b**
- **Not Equals: a != b**
- **Less than: a < b**
- **Less than or equal to: a <= b**
- **Greater than: a > b**
- **Greater than or equal to: a >= b**

```
a = 33
b = 200
if b > a:
  print("b is greater than a")
```

```
b is greater than a
```

An "if statement" is written by using the if keyword.

ELIF AND ELSE

- The elif keyword is pythons' way of saying "if the previous conditions were not true, then try this condition".

```
a = 200
b = 33
if b > a:
  print("b is greater than a")
elif a == b:
  print("a and b are equal")
else:
  print("a is greater than b")
```

```
a is greater than b
```

In this example a is equal to b, so the first condition is not true, but the elif condition is true, so we print to screen that "a and b are equal".

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

LOADING PYTHON LIBRARIES

In [ ]:

```
#Import Python Libraries
import numpy as np
import scipy as sp
import pandas as pd
import matplotlib as mpl
import seaborn as sns
```

Press `Shift+Enter` to execute the *jupyter* cell

Helpful Web Site in Python

The Deck is Stacked Against Developers

Machine learning is taught by academics, for academics.
That's why most material is so *dry* and *math-heavy*.

Developers need to know what works and how to use it.
We need *less math* and *more tutorials with working code*.

Welcome to Machine Learning Mastery!

Hi, I'm Jason Brownlee PhD and I help developers like you skip years ahead.

Discover how to get *better results*, *faster*.

Click the button below to get my free EBook and accelerate your next project
(*and access to my exclusive email course*).

Send Me the Free eBook!

Join over 150,000 practitioners who already have a head start.

END

Occam's razor
simpler explanations are more plausible and any unnecessary complexity should be shaved off