Answer the following questions (short answer):

1. What is the role of the hidden layer in Neural Networks ?
2. What is the difference between unsupervised learning and supervised learning ?
3. Give the values of the parameters β_0 *and* β_1 in the linear regression equation

Exercice 01

Draw the perceptron diagram and label each of its components.

Exercice 02

You are given the following dataset of students, where the goal is to predict whether a student passes an exam.

Student	Study Hours	Attendance (%)	Result
A	2	60	Fail
B	4	70	Fail
C	6	80	Pass
D	8	90	Pass
E	5	75	Pass

A new student F has a study hours = 5 and an attendance = 72%

Questions:

1. What type of learning algorithm is KNN?
2. If K = 3, identify the 3 nearest neighbors of student F.
3. What class (Pass or Fail) will KNN assign to student F?

Exercice 03

You are given a dataset of customers from a retail store. The goal is to group customers into clusters using K-means.

Customer	Age	Annual Income ($)	Spending Score	Gender
CU1	22	15,000	39	Male
CU2	789	NAN	81	Female
CU3	47	65,000	6	Male
CU4	52	70,000	77	Female
CU5	46	62,000	40	Female

- Clean and prepare the dataset (preprocessing).
- Apply the K-means algorithm with k = 2, where the initial cluster centers are C1 (0,0,0,0) and C2 (1,1,1,1).
- Give the customers assigned to each cluster (C1 or C2) after applying the K-means algorithm.

The exam model answer

Answers to the questions:

1. The role of the hidden layer: extract features
2. Supervised learning uses **labeled data**, where the correct output is known, to train a model. In contrast, unsupervised learning works with **unlabeled data**.
3.

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

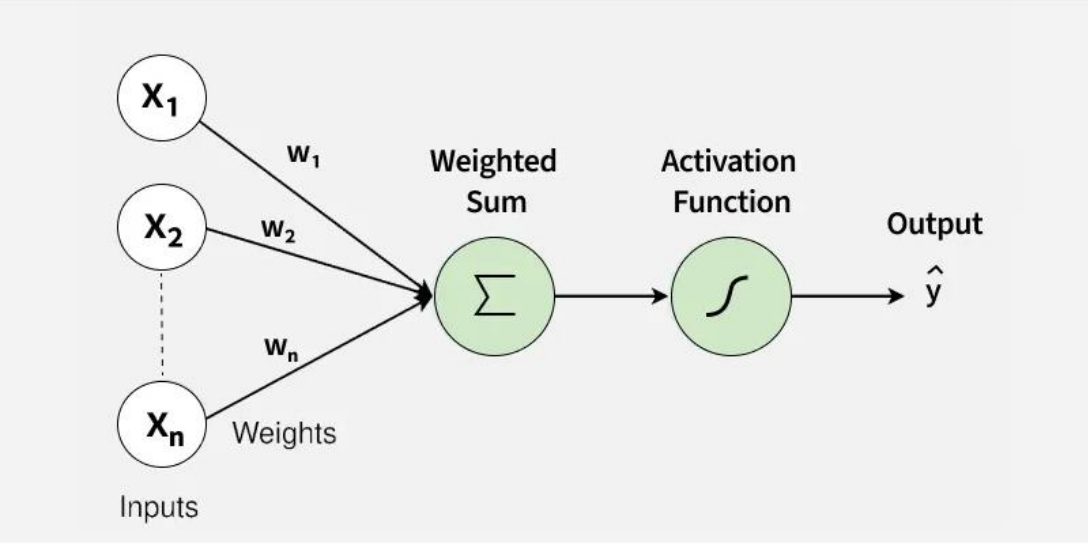
$$\beta_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Where:

- $\bar{x}$: mean of x
- $\bar{y}$: mean of y

Exercice 01

The perceptron Diagram



Exercice 02

1. KNN is a supervised learning algorithm
2. To find the 3 nearest neighbors to F, we need to calculate the Euclidian distance between F and the other students.

student	student	distance
F	A	$\sqrt{(5-2)^2+(72-60)^2}=12.36$
F	B	2.23
F	C	8.06
F	D	18.24
F	E	3

The 3 nearest neighbors are: B, C and E

3. The class of student F is **Pass** (because it is the most frequent class)

Exercise 03

1. Preprocessing:

- Replacing missing value:
  Age of CU2 = $\frac{22+47+52+46}{4} = 41.75$
- Replacing outlier:
  Annual income of CU2 = $\frac{15000+65000+70000+62000}{4} = 53000$
- Coding a categorical variable:
  Male =1 and female = 0
- Normalization: we use this formula to normalize each feature

$$x' = \frac{x - min}{max - min}$$

The resulting dataset is as follows:

Customer	Age	Annual Income ($)	Spending Score	Gender
CU1	0	0	0.44	1
CU2	0.65	0.69	1	0
CU3	0.83	0.90	0	1
CU4	1	1	0.94	0
CU5	0.80	0.85	0.45	0

2. K-means algorithm

- **Step 1**

Initial centroids C1 (0,0,0,0) and C2 (1,1,1,1).

Example:

distance between CU1 and C1 =

$\sqrt{(0-0)^2+(0-0)^2+(0-0.44)^2+(0-1)^2}$ = **1.09**

Customer	Distance to C1	Distance to C2
CU1	1.09	1.52
CU2	1.37	1.10
CU3	1.58	1.01
CU4	1.69	1.00
CU5	1.25	1.16

Cluster 1 = {CU1}
Cluster 2 = {CU2, CU3, CU4, CU5}
The new centroids are
C1 = (0, 0, 0.44, 1)
C2 = $(\frac{0.65+0.83+1+0.80}{4}, \frac{0.69+0.90+1+0.85}{4}, \frac{1+0+0.94+0.45}{4}, \frac{0+1+0+0}{4})$
= (0.82, 0.86, 0.59, 0.25)

- **Step 2**

The new centroids are: C1 = (0, 0, 0.44, 1) and C2 = (0.82, 0.86, 0.59, 0.25)
We now calculate the distances:

Customer	Distance to C1	Distance to C2
CU1	0	1.41
CU2	1.48	0.53
CU3	1.30	0.95
CU4	1.80	0.48
CU5	1.53	0.28

Cluster 1 = {CU1}
Cluster 2 = {CU2, CU3, CU4, CU5}
C1 = (0, 0, 0.44, 1)
C2 = $(\frac{0.65+0.83+1+0.80}{4}, \frac{0.69+0.90+1+0.85}{4}, \frac{1+0+0.94+0.45}{4}, \frac{0+1+0+0}{4})$
= (0.82, 0.86, 0.59, 0.25)
Since the centroids did not change, the algorithm stops.

3. the customers assigned to each cluster:
   Cluster 1 = {CU1}
   Cluster 2 = {CU2, CU3, CU4, CU5}