

TD Series 2: Supervised learning

Exercise 1 :

A dataset containing medical records of patients, including factors like age, blood pressure, cholesterol levels, and family history and smoking habits. Your task is to build a model that predicts whether a patient is at risk of developing heart disease.

1. What is the type of this supervised learning problem?
2. List the features that could be used to predict heart disease risk.
3. Define the output labels.
4. What type of data will be used to train the model?
5. How will the model's performance be evaluated?
6. What additional features could be included in the dataset to enhance the model's ability to predict heart disease risk?

Exercise 2 :

The dataset below describes the employees' years of experience and their corresponding salaries. We use **Simple Linear Regression** to predict the salary of an employee based on their years of experience.

Experience (Years)	Salary (\$)
1	40,000
2	45,000
3	50,000
4	55,000
5	60,000
6	65,000

1. Visualize the relationship between experience (X) and salary (Y) in a plot.
2. Use the least squares method to fit a line and find the equation $y=b+mx$.
3. Predict the salary for an employee with 7 years of experience.
4. Compute **R-squared** and **MSE** to assess how well the model fits the data.

Exercise 3:

A retail company wants to predict whether a customer will make a purchase based on their characteristics. The company has collected data about customer information and construct a dataset. The dataset includes the following features: Age: The age of the customer (in years), Income: The annual income of the customer (in dollars). Previous Purchase: Whether the customer has made a previous purchase (Yes or No). Marital Status: Whether the customer is married or single (Married/Single).

Customer id	Age	Income	Previous purchase	Marital statue	Purchase
1	< 30	< 45000	Yes	Single	Yes
2	≥ 30	≥ 45000	Yes	Married	Yes
3	≥ 30	≥ 45000	No	Married	No
4	< 30	< 45000	No	Single	No
5	≥ 30	≥ 45000	Yes	Married	Yes
6	< 30	< 45000	No	Single	No
7	≥ 30	≥ 45000	Yes	Married	Yes
8	≥ 30	≥ 45000	No	Married	Yes
9	< 30	< 45000	No	Single	No
10	≥ 30	≥ 45000	Yes	Married	Yes

1. Define the target variable.
2. Explain the decision tree objectives .
3. Explain the decision tree steps.
4. Build the decision tree corresponding to this dataset using ID3 algorithm.
5. Given the following information about a customer:

Customer id	Age	income	Previous purchase	Marital status
11	30	55,000	Yes	Single

What would the Decision Tree model predict? Will the customer make a purchase (Yes or No)?