ML

LECTURE-2

BY
Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

Regression

- **Definition**
- In machine learning, a regression problem is the problem of **predicting the value of a numeric variable based on observed values of the variable.** The **value of the output variable may be a number, such as an integer or a floating point value.**
- These are often quantities, such as amounts and sizes. **The input variables may be discrete or real-valued.**
- **Example**
- Consider the data on car prices given in Table 1.2.

Price (US$)	Age (years)	Distance (KM)	Weight (pounds)
13500	23	46986	1165
13750	23	72937	1165
13950	24	41711	1165
14950	26	48000	1165
13750	30	38500	1170
12950	32	61000	1170
16900	27	94612	1245
18600	30	75889	1245
21500	27	19700	1185
12950	23	71138	1105

Table 1.2: Prices of used cars: example data for regression

- Suppose we are required to estimate the price of a car aged 25 years with distance 53240 KM and weight 1200 pounds. It is an ex. of a regression problem beause we have to predict the value of the numeric variable “Price”.

Regression

- **General approach**
- Let x denote the set of input variables and y the output variable.
- In machine learning, the general approach to regression is to assume a model, that is, some mathematical relation between x and y, involving some parameters say, θ, in the following form:

$$y = f(x, \theta)$$

- The function $f(x, \theta)$ is called the **regression function**.
- The **machine learning algorithm optimizes the parameters in the set θ such that the approximation error is minimized;** that is, the estimates of the values of the dependent variable y are as close as possible to the correct values given in the training set.

- **Example**
- For example, if the input variables are “Age”, “Distance” and “Weight” and the output variable is “Price”, the model may be

$$y = f(x, \theta)$$

- Price = $a_0 + a_1 \times (\text{Age}) + a_2 \times (\text{Distance}) + a_3 \times (\text{Weight})$
- where x = (Age, Distance, Weight) denotes the the set of input variables and $\theta = (a_0, a_1, a_2, a_3)$ denotes the set of parameters of the model.

Regression

- **Different regression models**
- There are various types of regression techniques available to make predictions. These techniques mostly differ in **three aspects, namely, the number and type of independent variables, the type of dependent variables and the shape of regression line.** Some of these are listed below.
- **Simple linear regression: There is only one continuous independent variable x** and the assumed relation between the independent variable and the dependent variable y is

$$y = a + bx.$$

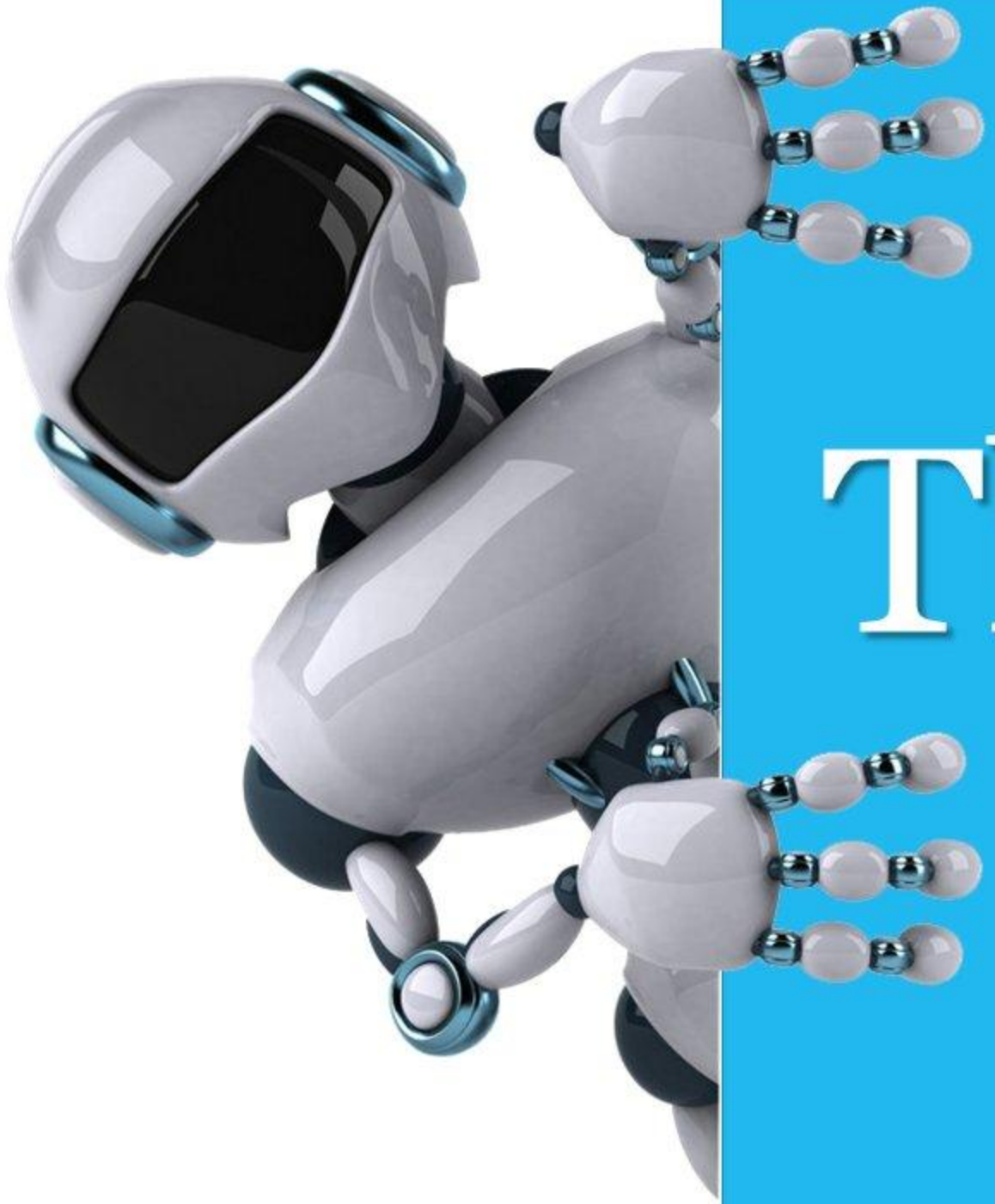
- **Multivariate (Multiple) linear regression: There are more than one independent variable, say $x_1, \ldots, x_n$**, and the assumed relation between the independent variables and the dependent variable is

$$y = a_0 + a_1x_1 + \cdots + a_nx_n.$$

- **Polynomial regression: There is only one continuous independent variable x** and the assumed model is

$$y = a_0 + a_1x + a_2x^{2\cdots} + a_nx^n, \text{ (for some positive integer } n > 1)$$

- **Logistic regression:** The **dependent variable is binary (0/1, True/False, Yes/No)** in nature. Even though the output is a binary variable, what is being sought is a **probability function which may take any value from 0 to 1.**



Thank you