CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

National University of Computer and Emerging Sciences,

Faisalabad, Pakistan.

REGRESSION VS CLASSIFICATION

Classification

Classification problem

Features | Label

#	Height (inches)	Weight (kgs)	B.P. Sys	B.P. Dia	Heart disease
1	62	70	120	80	No
2	72	90	110	70	No
3	74	80	130	70	No
4	65	120	150	90	Yes
5	67	100	140	85	Yes
6	64	110	130	90	No
7	69	150	170	100	Yes
8	66	125	145	90	?
9	74	67	110	60	?

Feature vector (4-dimensional)

Label vector

Training Data

Test Data

Regression

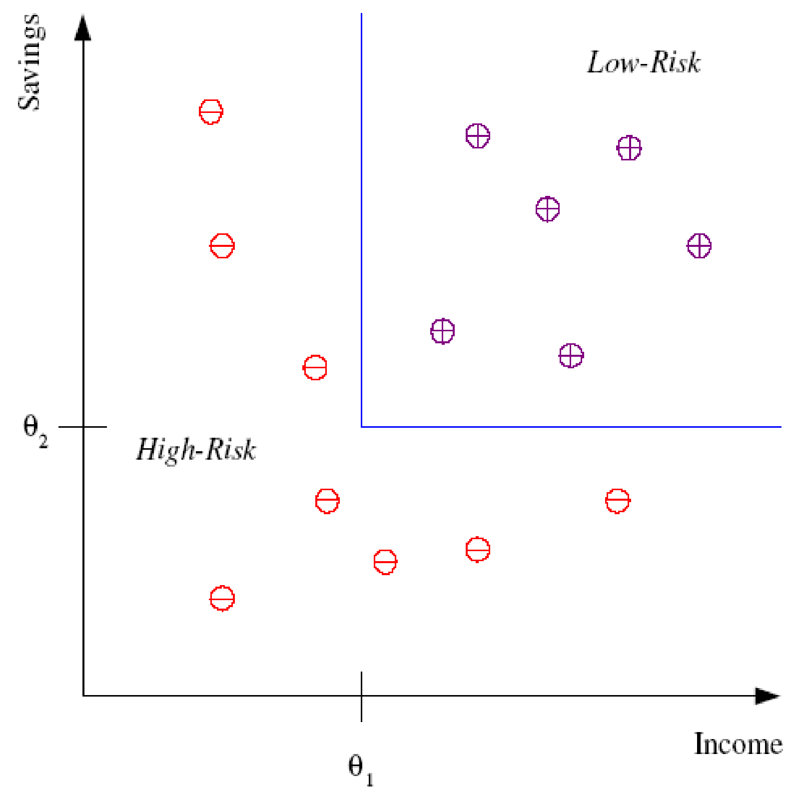
Regression problem

#	Height (inches)	Weight (kgs)	B.P. Sys	B.P. Dia	Cholesterol Level
1	62	70	120	80	150
2	72	90	110	70	160
3	74	80	130	70	130
4	65	120	150	90	200
5	67	100	140	85	190
6	64	110	130	90	130
7	69	150	170	100	250
8	66	125	145	90	?
9	74	67	110	60	?

Classification

Classification

Predict discrete-valued output



Differentiating between low-risk and high-risk customers from their *income* and *savings*

Discriminant Model:

IF income > θ1 AND savings > θ2

THEN low-risk

ELSE high-risk

Regression

Regression

- Predict real-valued output

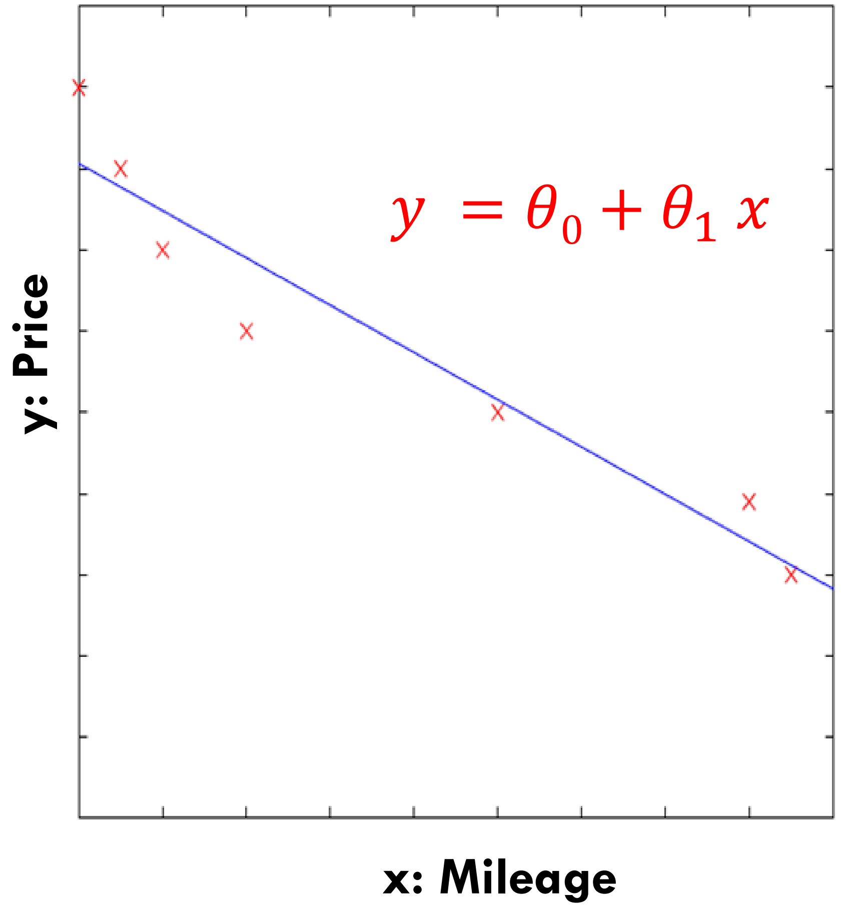
Examples: Price of a Used Car

- **Inputs:** are the car attributes—brand, year, engine capacity, mileage, and other information—that may affect a car's price.
- **Output:** is the price of the car.
- Such problems where the output is a **number,** are *regression* problems.

Regression Example

- **X**: car attributes (input variables)
- **Y**: the price of the car (target/output variables)
- Learn the program that fits the function to training examples to learn **Y** as the function of **X**.

$$\boldsymbol{y = \theta_0 + \theta_1 x}$$



Error Measure

Classification

- Y is discrete, a (small) finite, unordered set of classes

$$error\big(h(x), f(x)\big) = 0 \text{ if } h(x) = f(x) \text{ else } 1$$

0-1 Loss Error

Regression

- Y is continuous, a numeric set (typically real numbers)

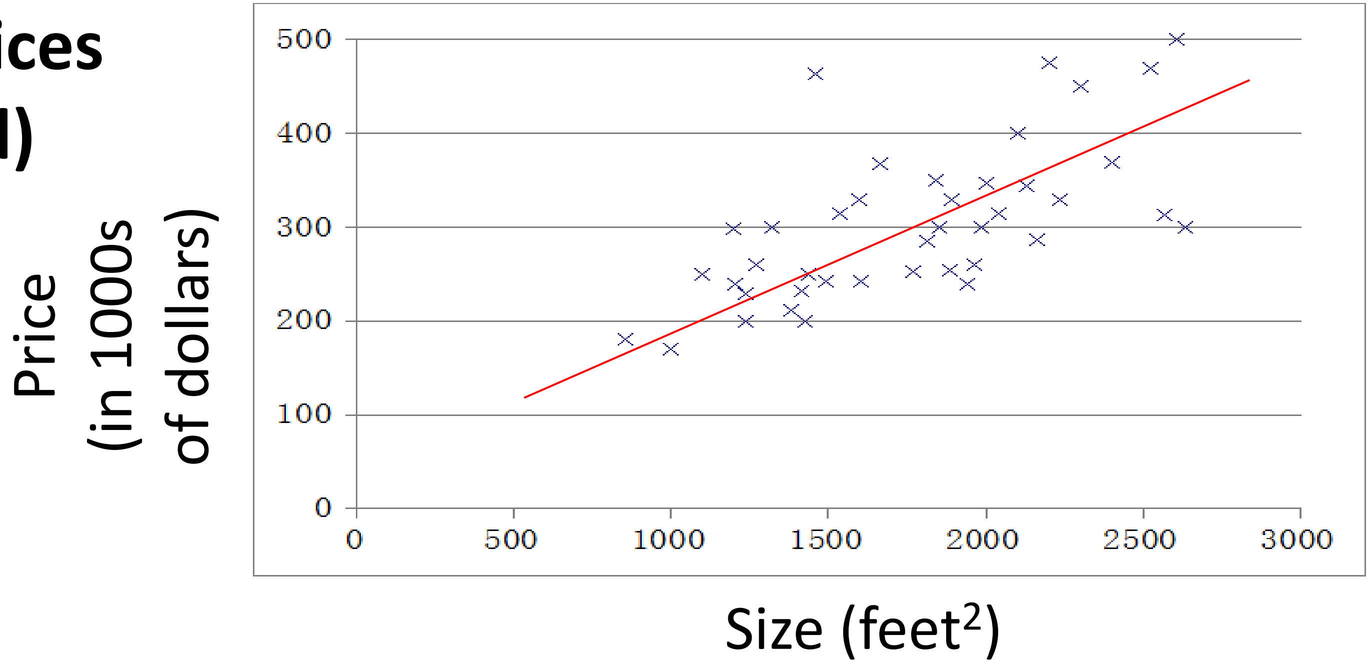
$$error\big(h(x), f(x)\big) = (h(x) - f(x))^2$$

Squared Error

LINEAR REGRESSION

Linear Regression with one Variable

Housing Prices (Portland)



Supervised Learning

Given the "right answer" for each example in the data.

Regression Problem

Predict the *real-valued* ***continuous*** output

Regression Example

Training set of housing prices

Size in feet^2 (x)	**Price ($) in 1000's (y)**
2104	460
1416	232
1534	315
852	178
…	…

Notation:

m = Number of training examples

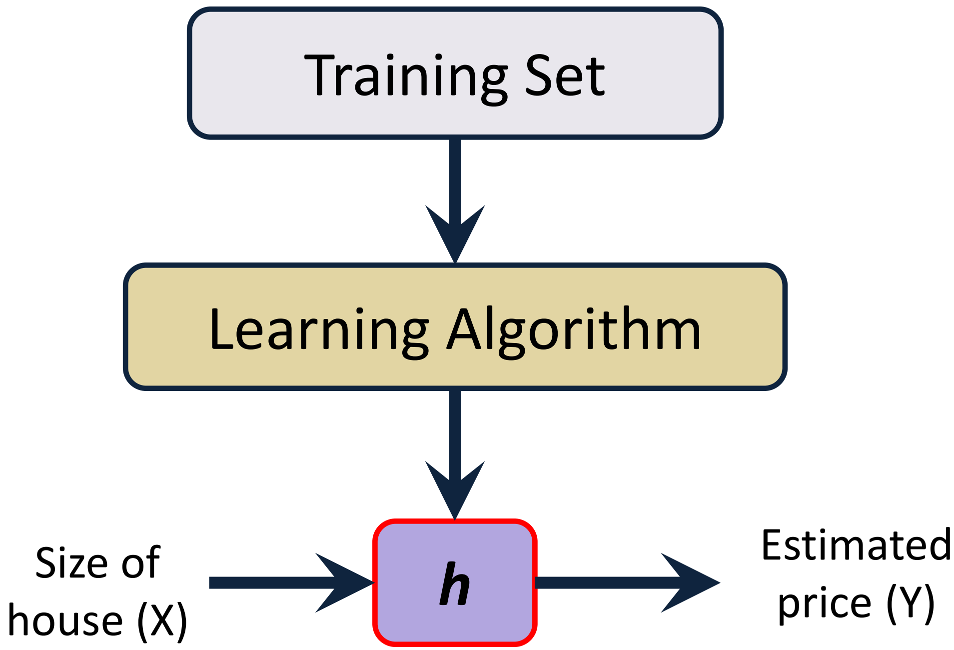
x's = "input" variable / features

y's = "output" variable / "target" variable

One Training example (x, y)

i^{th} training example $(x^{(i)}, y^{(i)})$

Regression



Question : How to describe *h*?

$$h: X \rightarrow Y$$

Regression Example

Training set of housing prices

Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

θ_i's: Parameters

How to choose θ_i's ?

Regression

How to choose these parameters, θ (regression coefficient)?

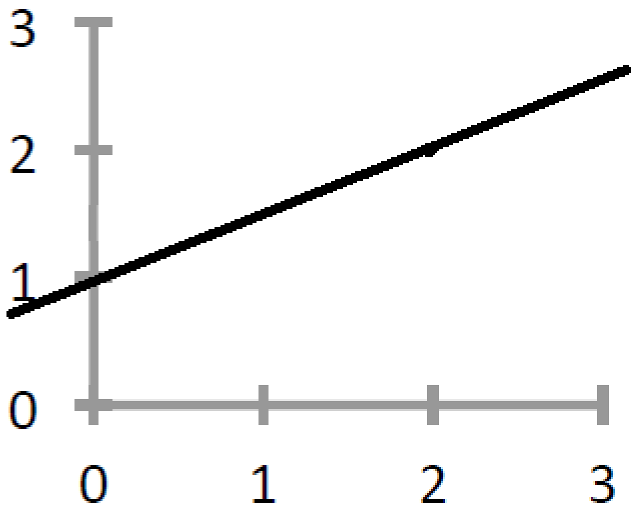
- The standard approach is the **least square method,** through which parameters are minimized.

- The machine learning program optimizes the parameters, θ, such that the approximation error is minimized.

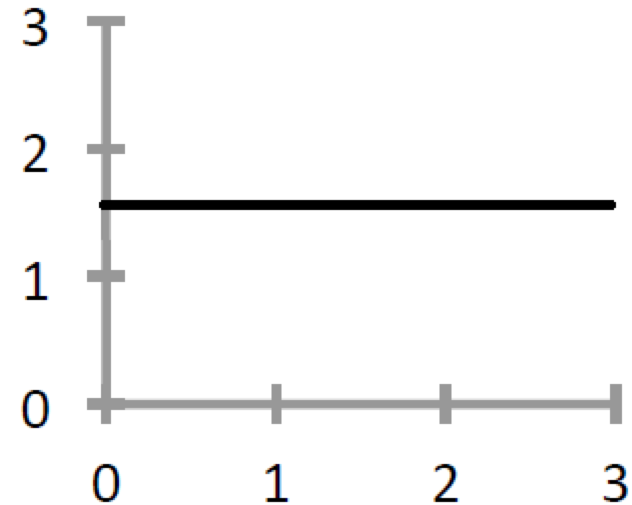
Regression

$$y = \theta_0 + \theta_1 x$$

$$\theta = \frac{change\ in\ Y}{change\ in\ X}$$



$$\theta_1 = 0.5, \theta_0 = 1$$

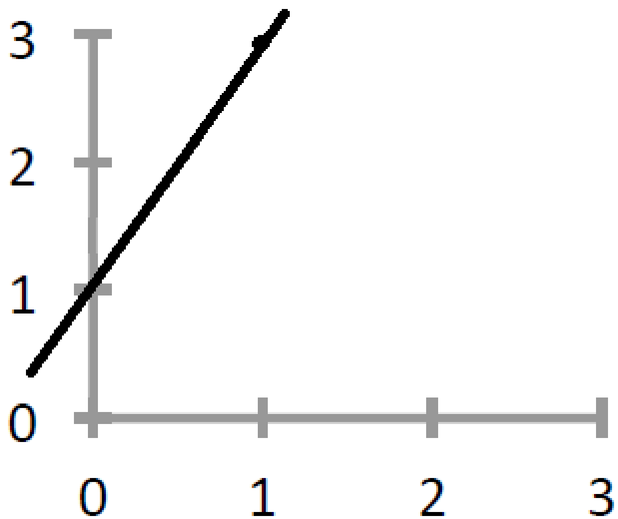


$$\theta_1 = 0, \theta_0 = 1.5$$

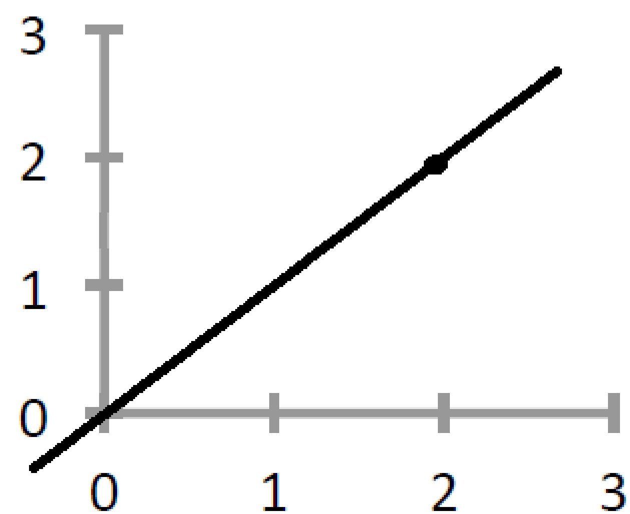
Regression

$$y = \theta_1 x + \theta_0$$

$$\theta = \frac{change\ in\ Y}{change\ in\ X}$$

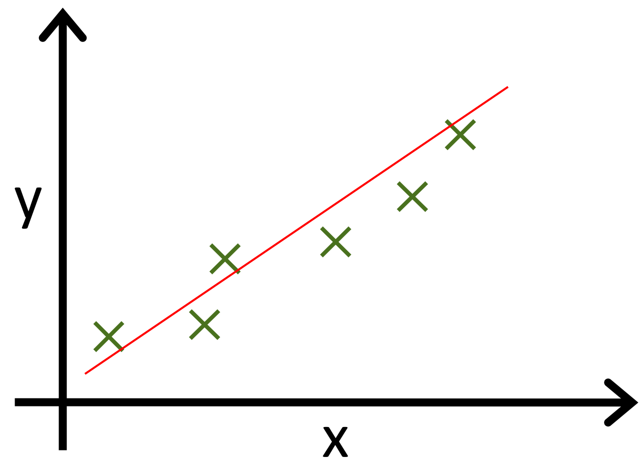


$$\theta_1 = 2, \theta_0 = 1$$



$$\theta_1 = 1, \theta_0 = 0$$

Regression



Idea: Choose θ_0, θ_1 so that
$h_\theta(x)$ is close to y for our
training examples (x, y)

Cost Function

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

Simplified:

$$h_\theta(x) = \theta_1 x$$

$$\theta_1$$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$$

Acknowledgement

Tom Mitchel, Russel & Norvig, Andrew Ng, Alpydin & Ch. Eick.