innovate achieve lead

BITS Pilani

Machine Learning(IS ZC464)

Session 2: Training and Testing in Learning systems

Prediction

X	Y
1	1
5	5
2	2
4	4
3	3

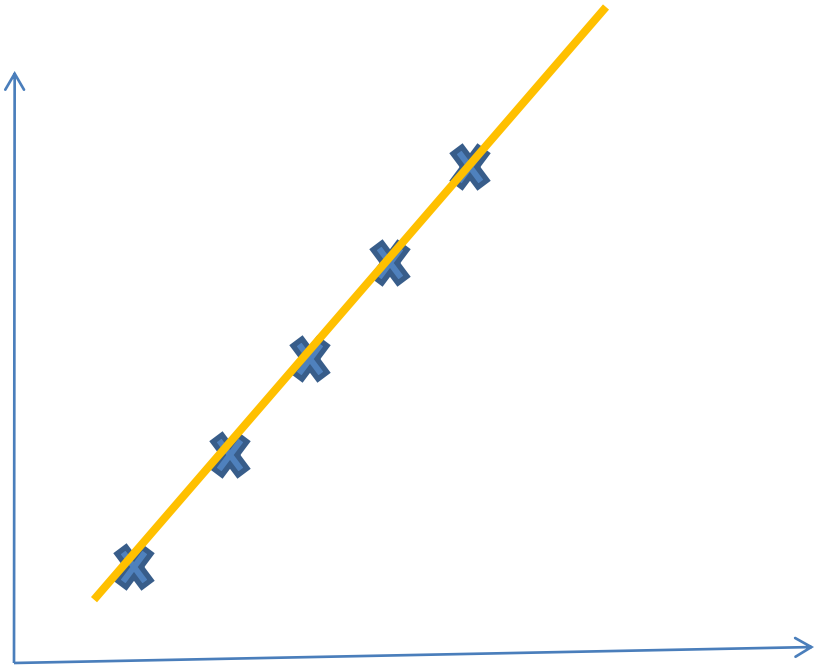
- Recall Learning: A machine with learning capability can predict about the new situation (seen or unseen) using its past experience.
- Prediction:

  Given values of x and y

  Predict value of y for x = 71
- Prediction is based on learning of the relationship between x and y
- Training data is the collection of (x,y) pairs
- Testing data is simply value of x for which value of y is required to be predicted.

Learning of a function from given sample data

Straight Line



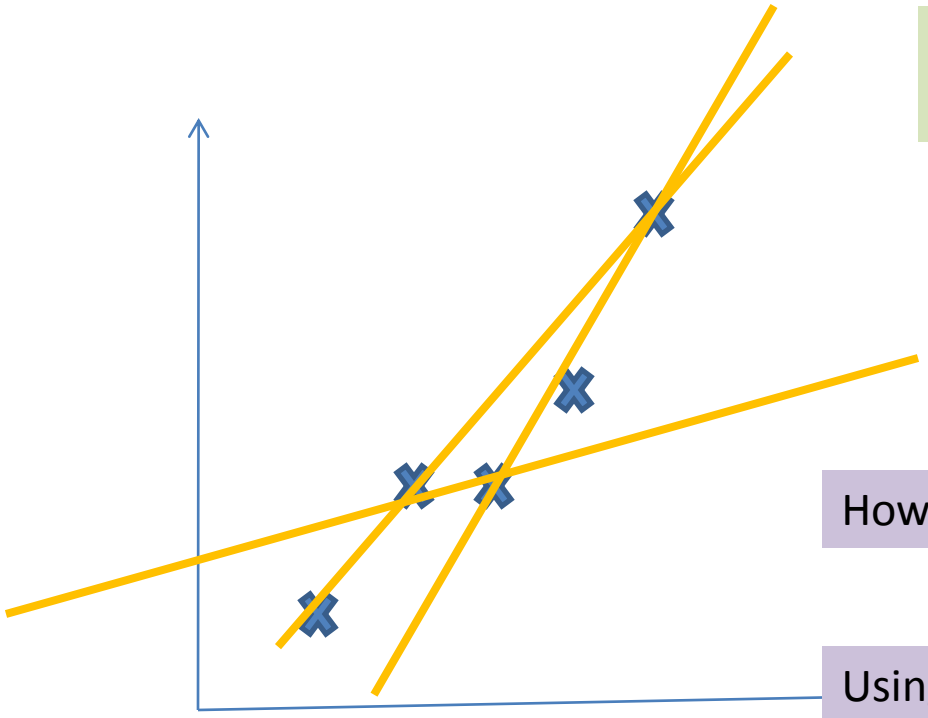
What did the system learn?

X	Y
1	1
5	5
2	2
4	3
3	2

- Y = f(x)
- Y = x
- What is its generalization ability?
- Most accurate or we can say 100%
- What if the data to train the system changes slightly? The machine can be still made to learn.

Learning of a function from given sample data-straight line learning

Straight Line



Which line fits the best?

Line is represented by parameters of slope and intercept

How?

Machine must learn on its own-which is the best fit

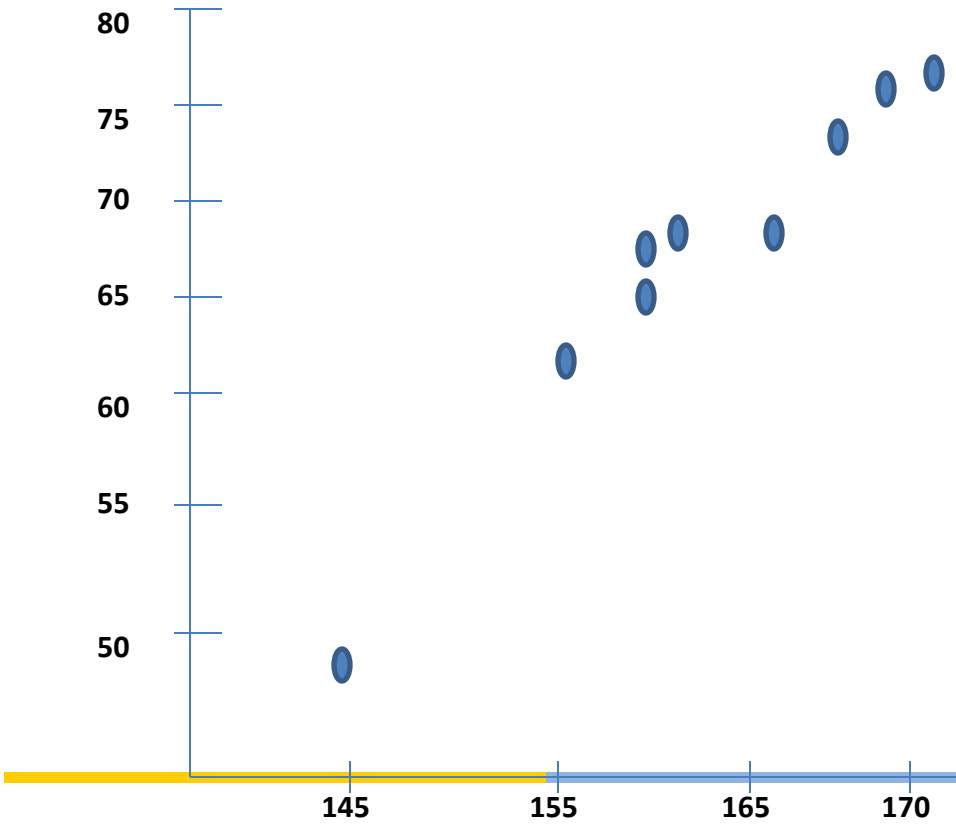
Using the data – known as training data i.e. (x,y) pair

Understanding ERROR

- Consider an example of using height and weight

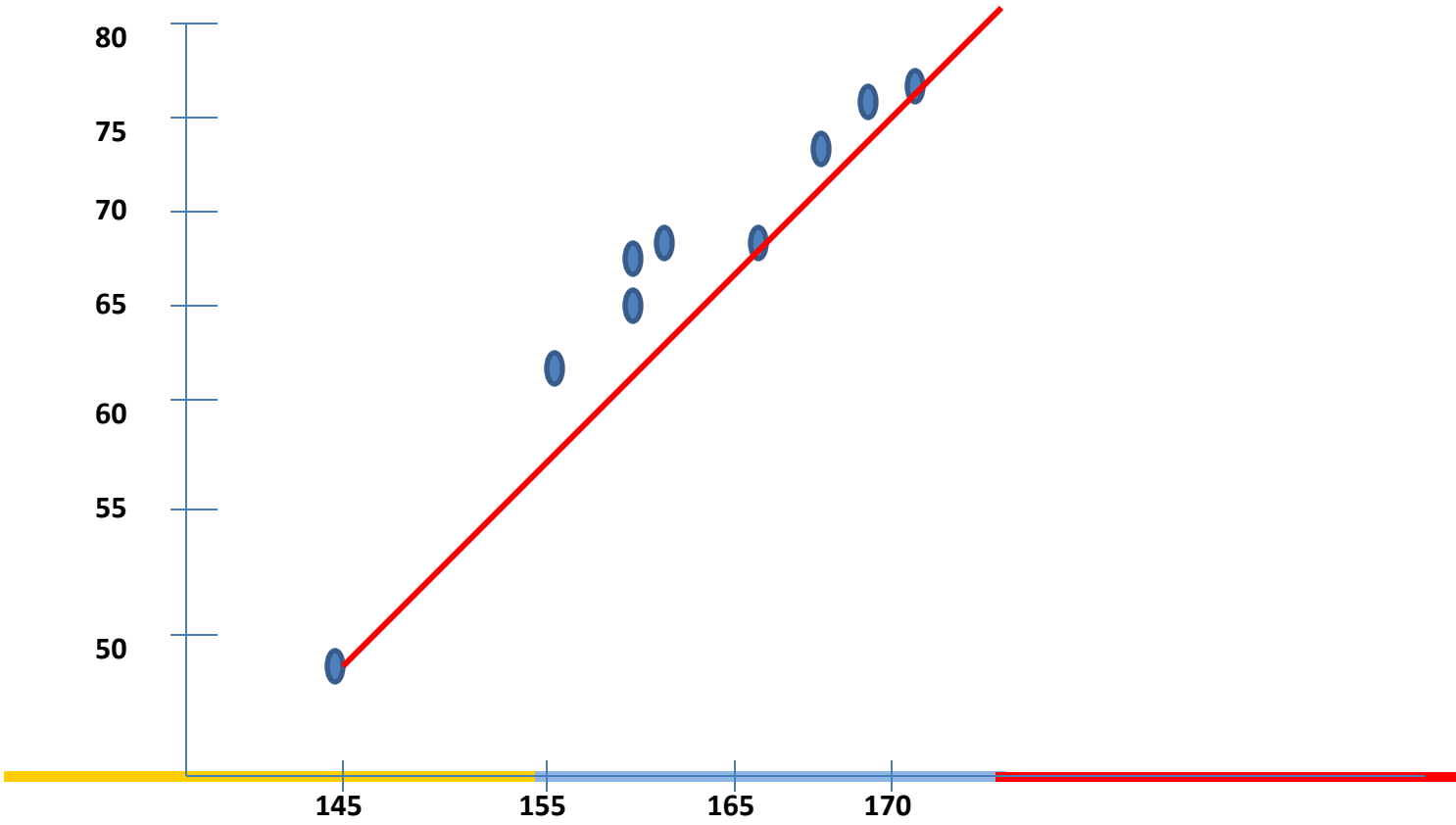
Height (in cm)	Weight (in Kg)
145	48
165	68
155	62
160	65
170	75
163	67
171	76
167	72
159	65

Understanding ERROR



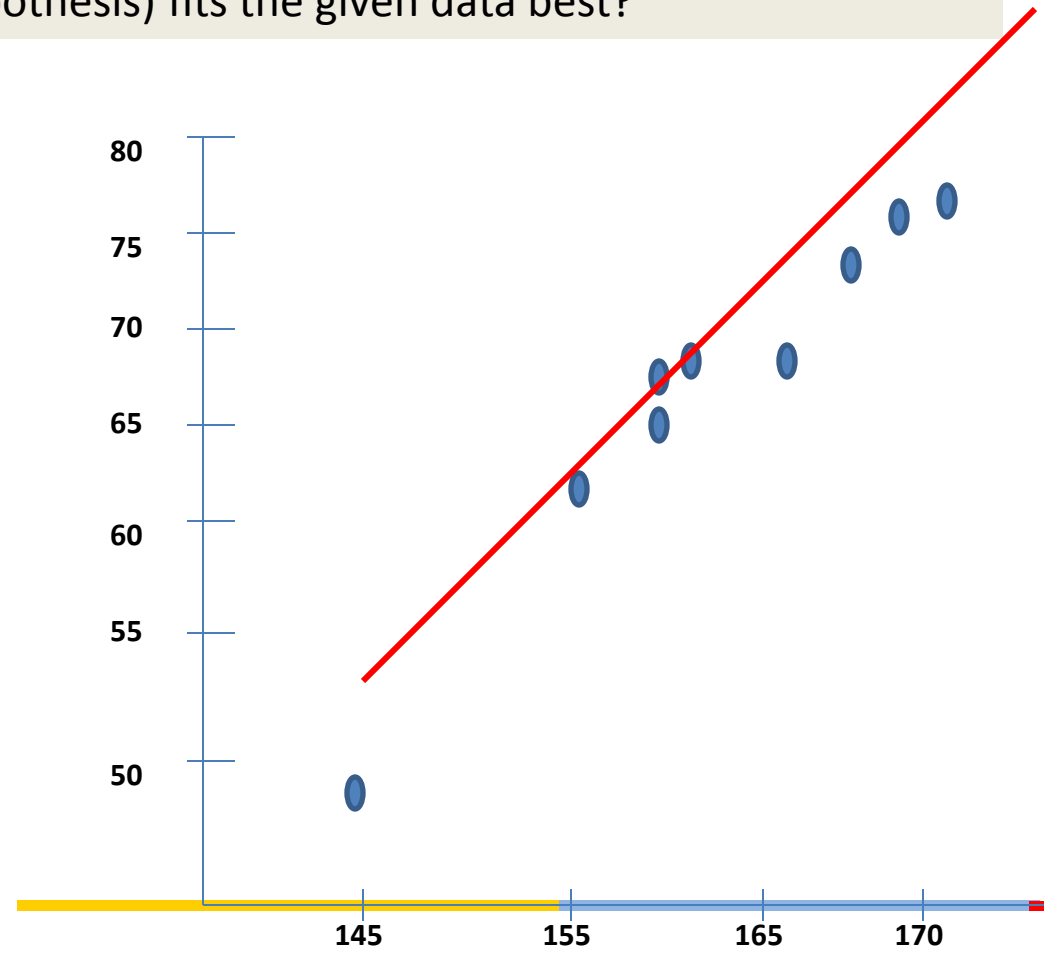
Understanding ERROR

Which line (hypothesis) fits the given data best?



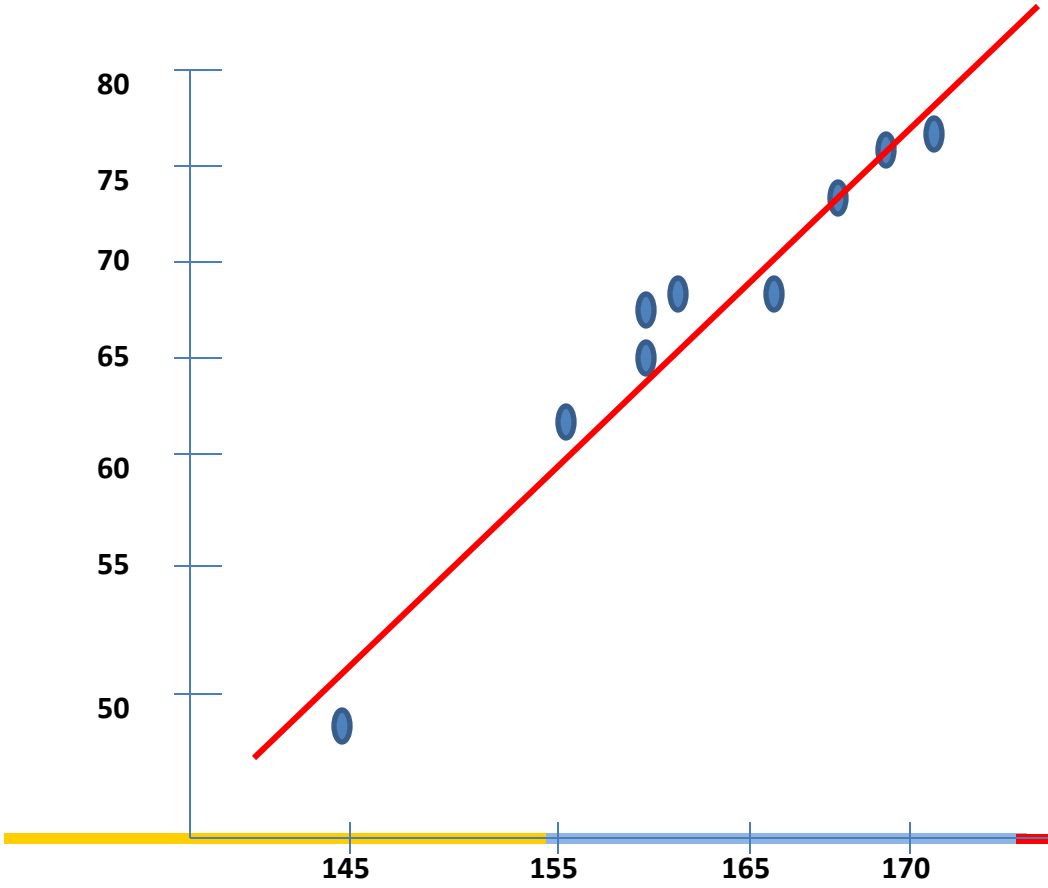
Understanding ERROR

Which line(hypothesis) fits the given data best?



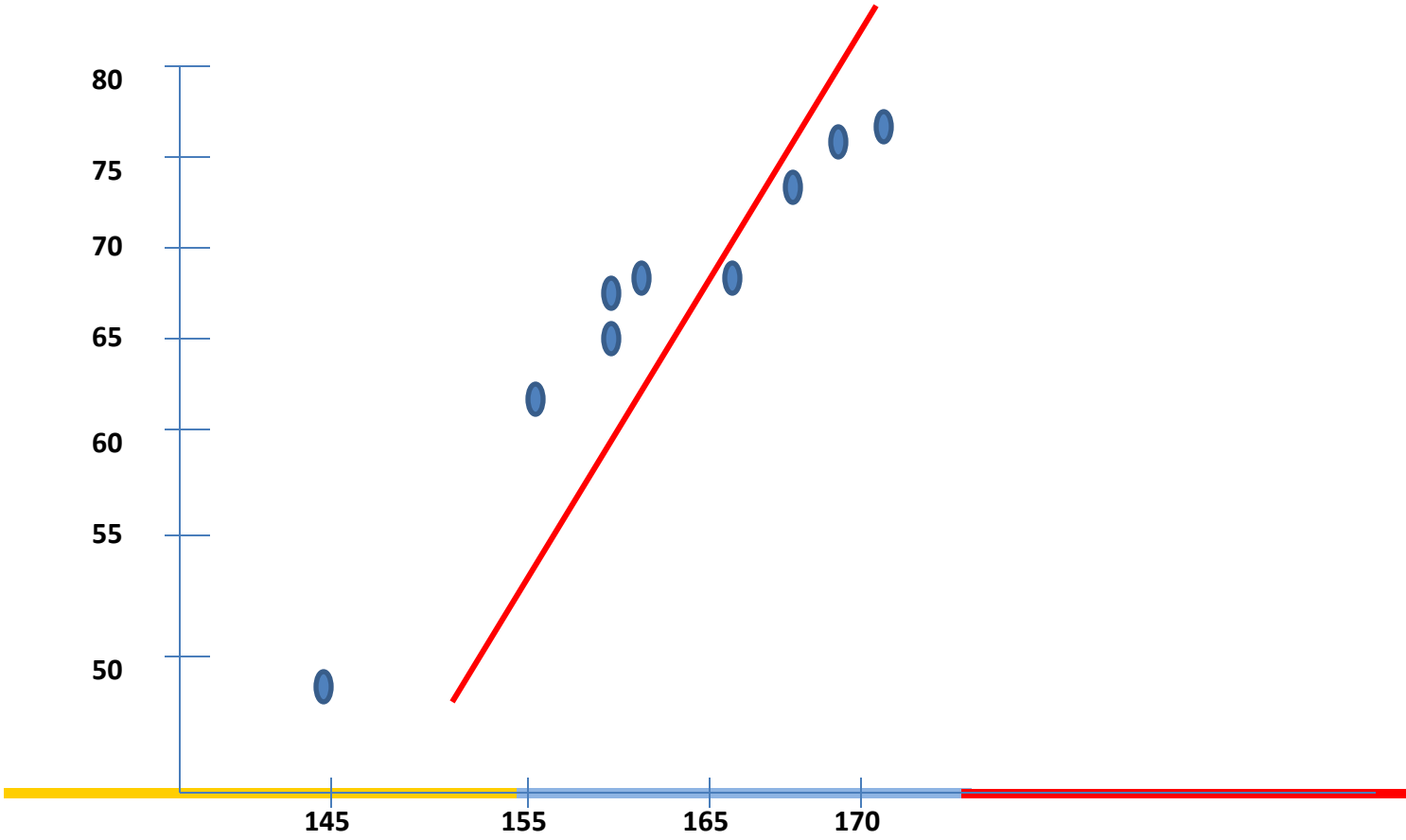
Understanding ERROR

Which line(hypothesis) fits the given data best?



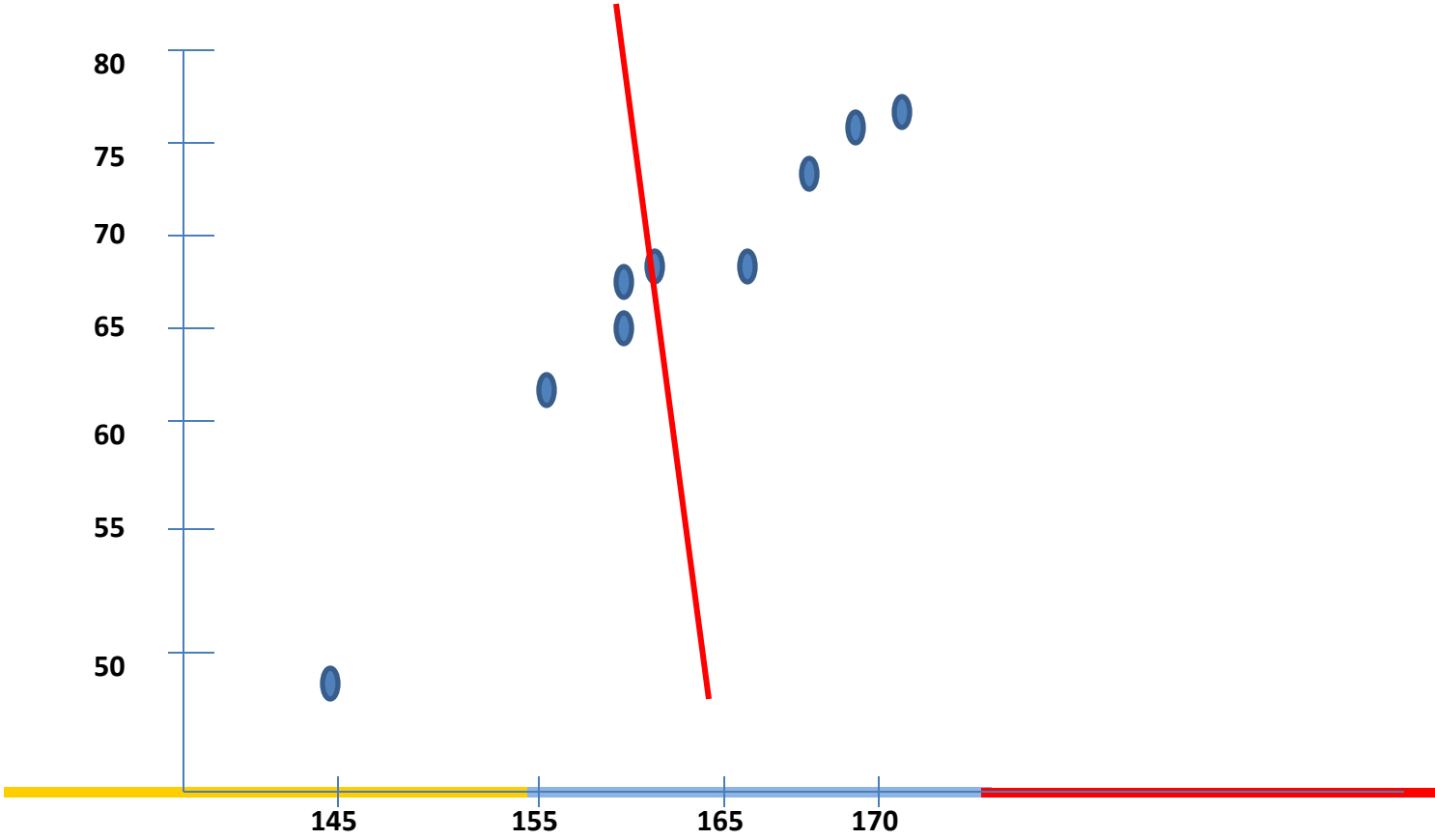
Understanding ERROR

Which line(hypothesis) fits the given data best?



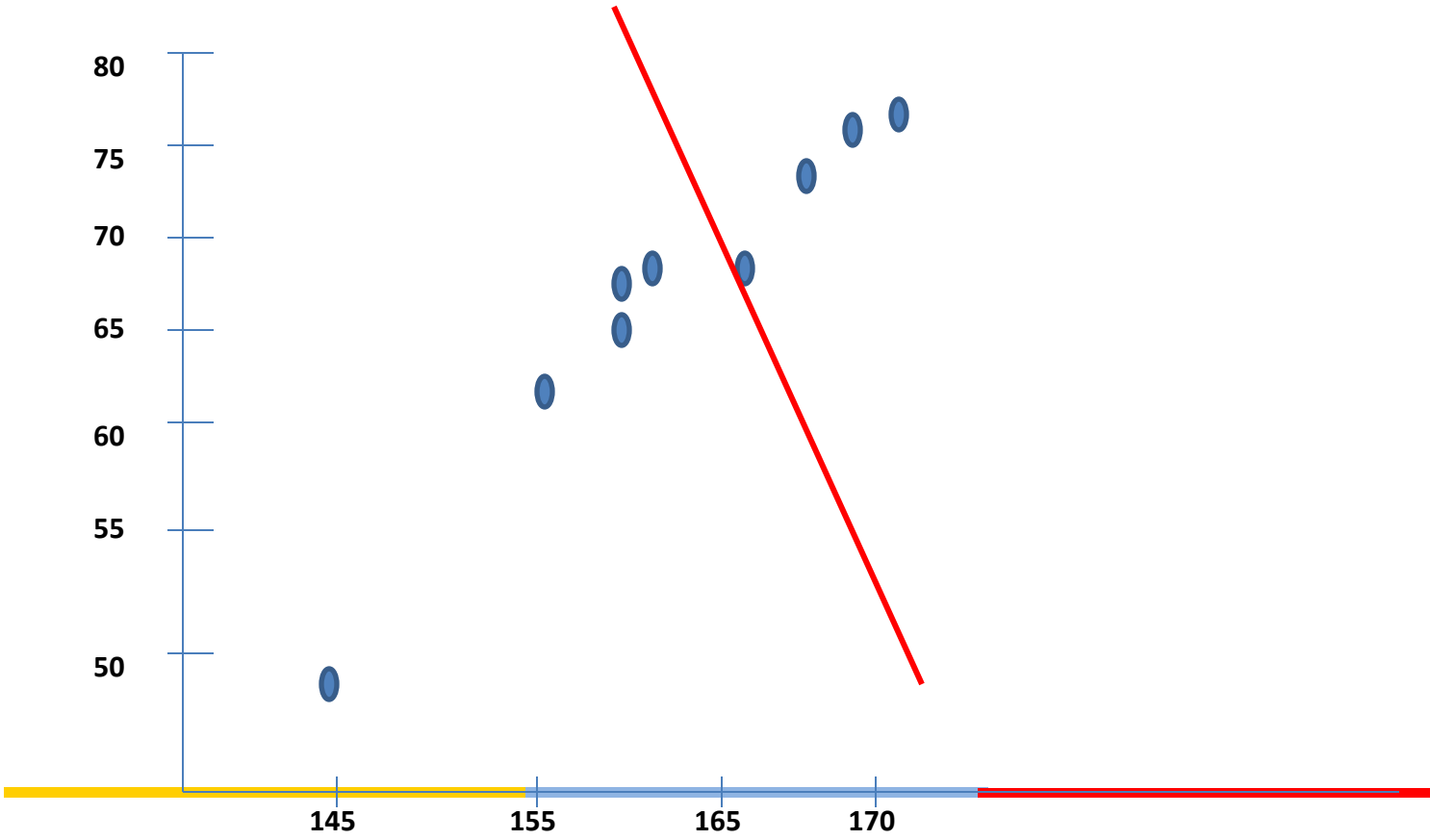
Understanding ERROR

Which line(hypothesis) fits the given data best?



Understanding ERROR

Which line(hypothesis) fits the given data best?

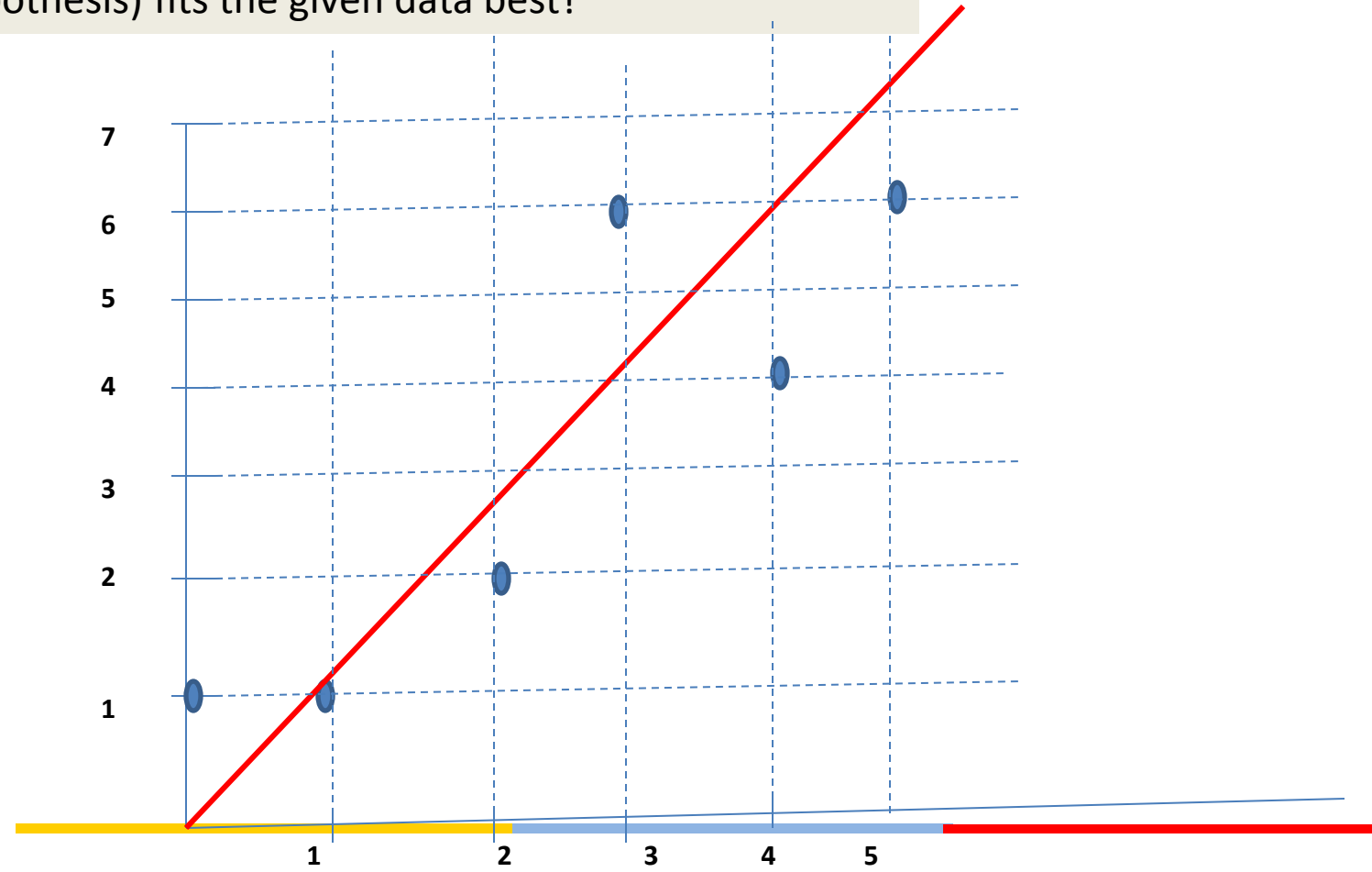


In 2D space the line parameters are two

- Slope and intercept
- Can be called as w_1 and w_2
- In order to find a line that best fits the given data, we must find w1 and w2 in such a way that the sum of the squared error is minimum

A simple example to understand ERROR

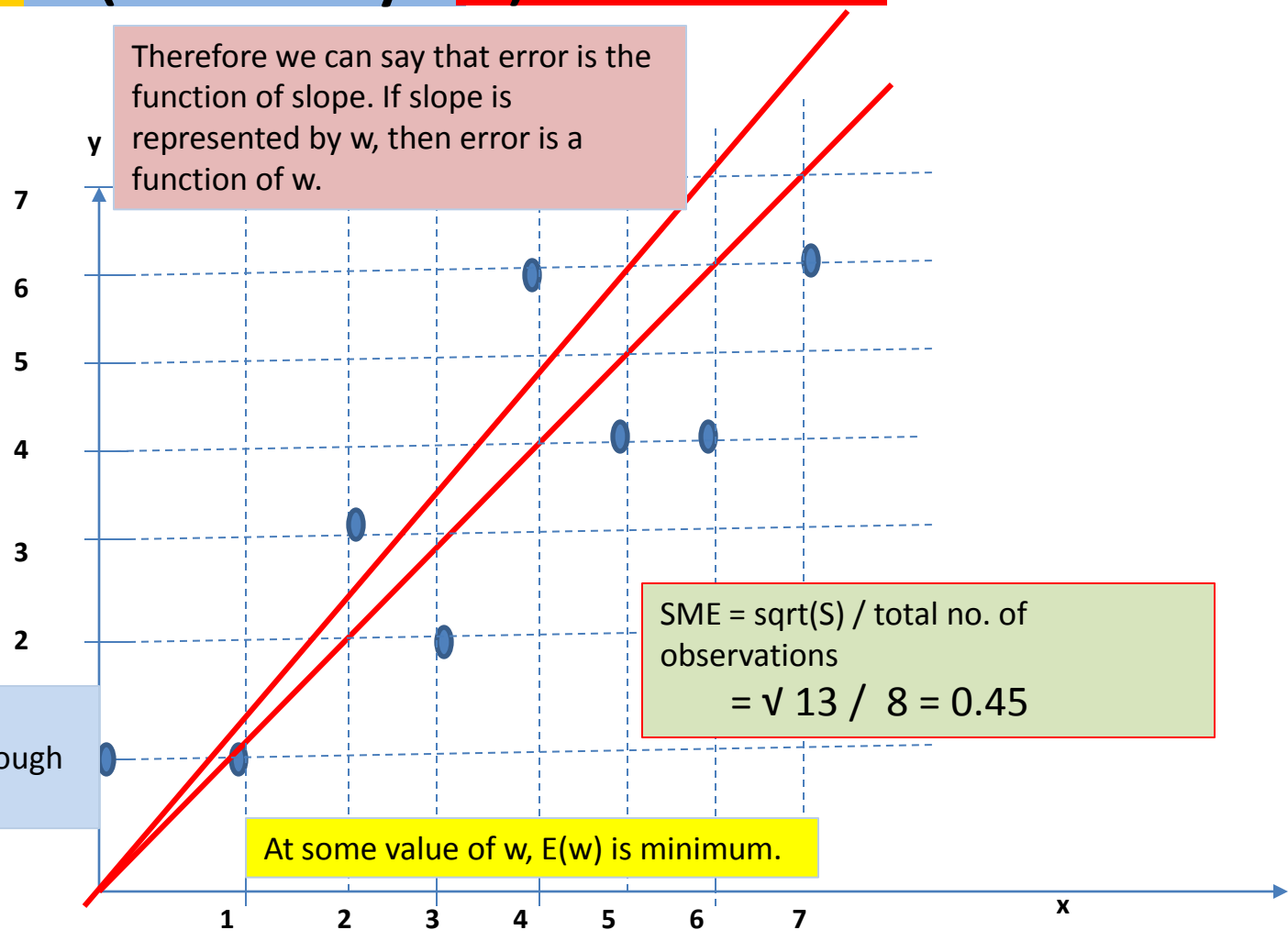
Which line(hypothesis) fits the given data best?



Compute the Squared Mean Error (line is y=x)

Sum of squares
(S) = 1*1
+ 0
+ 1*1
+1*1
+2*2
+1*1
+2*2
+1*1
=13

Therefore we can say that error is the function of slope. If slope is represented by w, then error is a function of w.

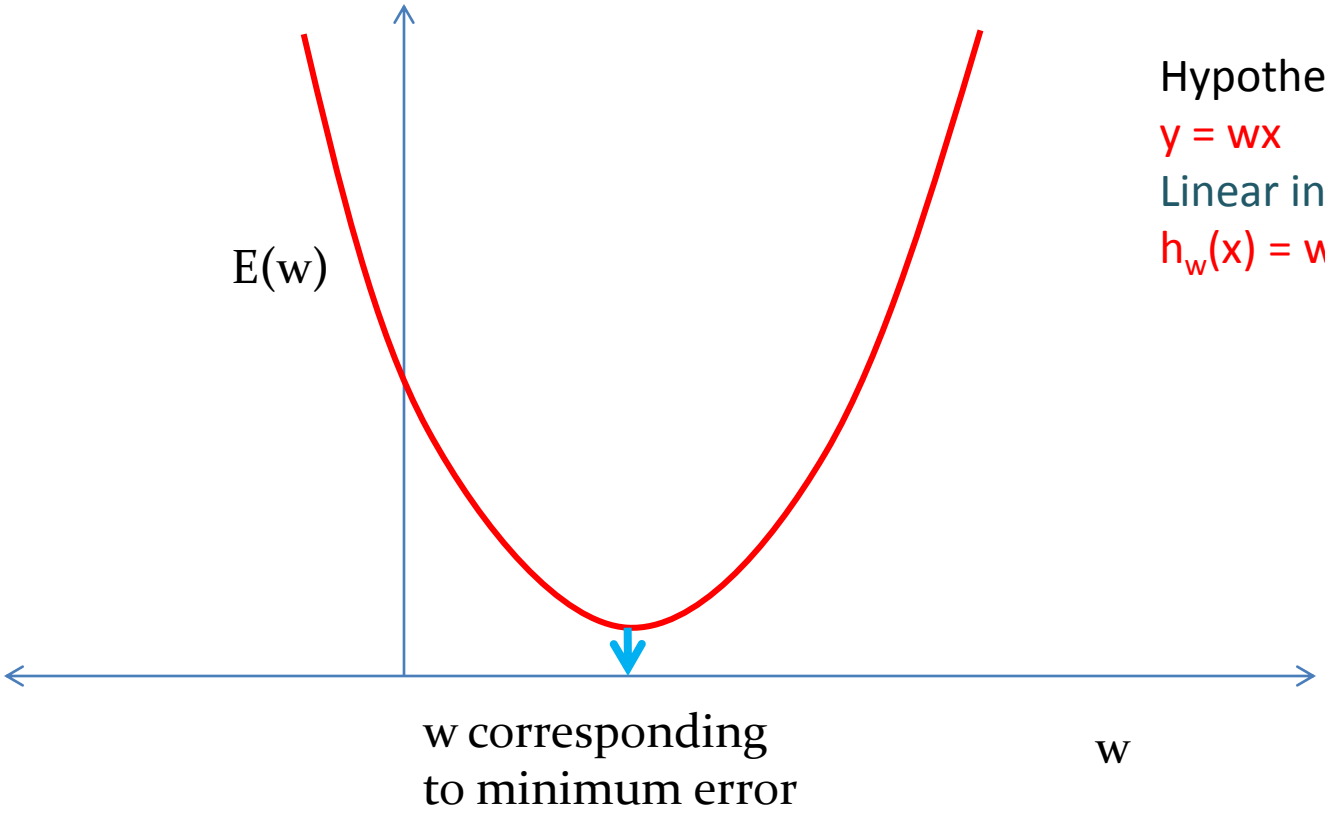


SME = sqrt(S) / total no. of observations
= √ 13 / 8 = 0.45

Error will be different if the line's slope is different (line passes through origin)

At some value of w, E(w) is minimum.

Plotting error when y=f(x)



Hypothesis function

$y = wx$

Linear in one variable

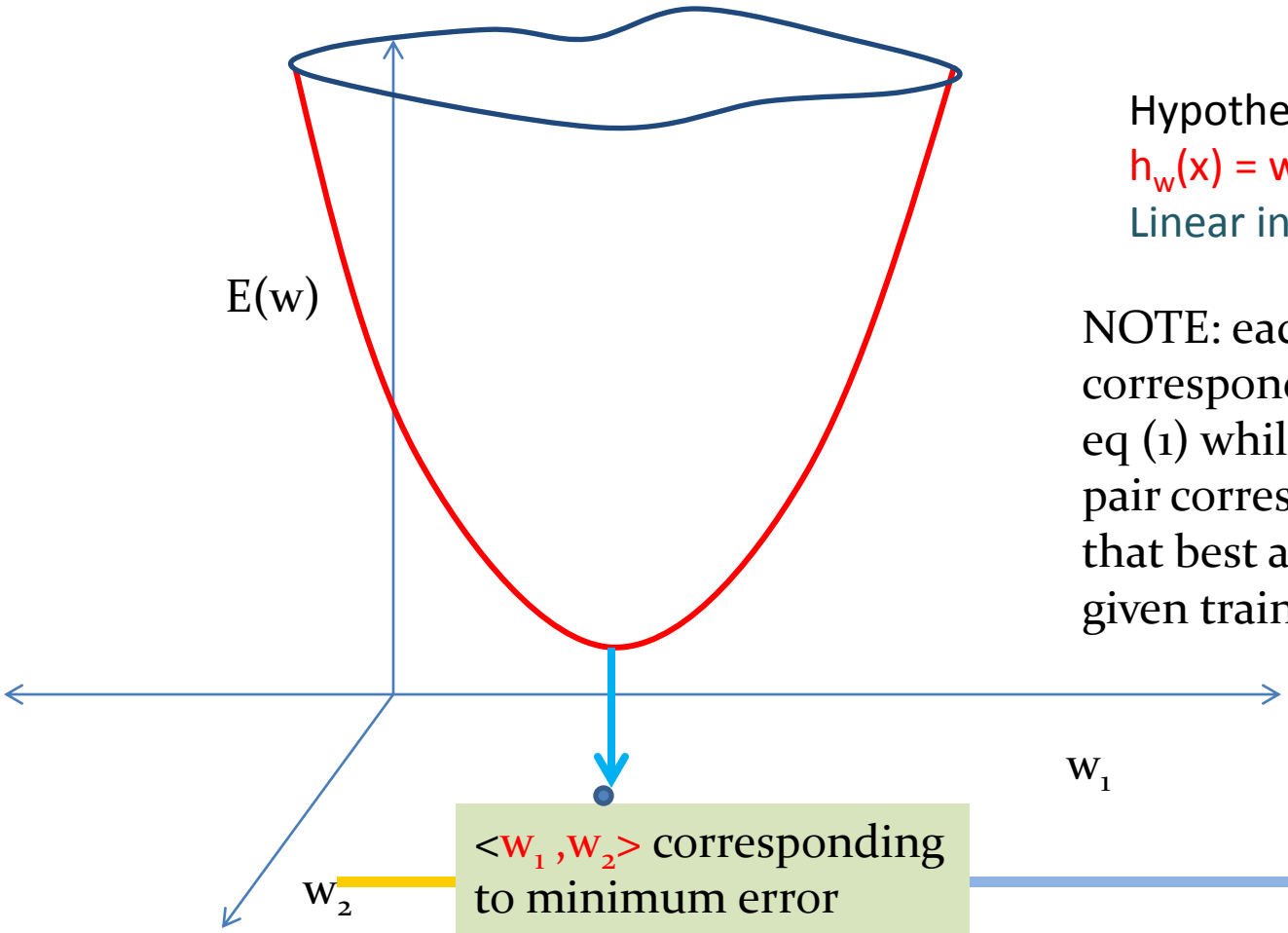
$h_w(x) = wx$

Understanding of the error surface

- Consider m observations $<x^1,y^1>$, $<x^2,y^2>$, ….$<x^m,y^m>$.
- An hypothesis $h_w(x)$ that approximates the function that fits best to the given values of y
- There is likely to be some error corresponding to each observation (say i).
- The magnitude of such error is $y^i - h_w(x^i)$
- Objective is to find such w that minimizes the sum of squares of errors

$$E_{min}(w) = \text{Minimize}_w \sum_i (y^i - h_w(x^i))^2$$

Plotting error when $y=f(x_1,x_2)$



Hypothesis function

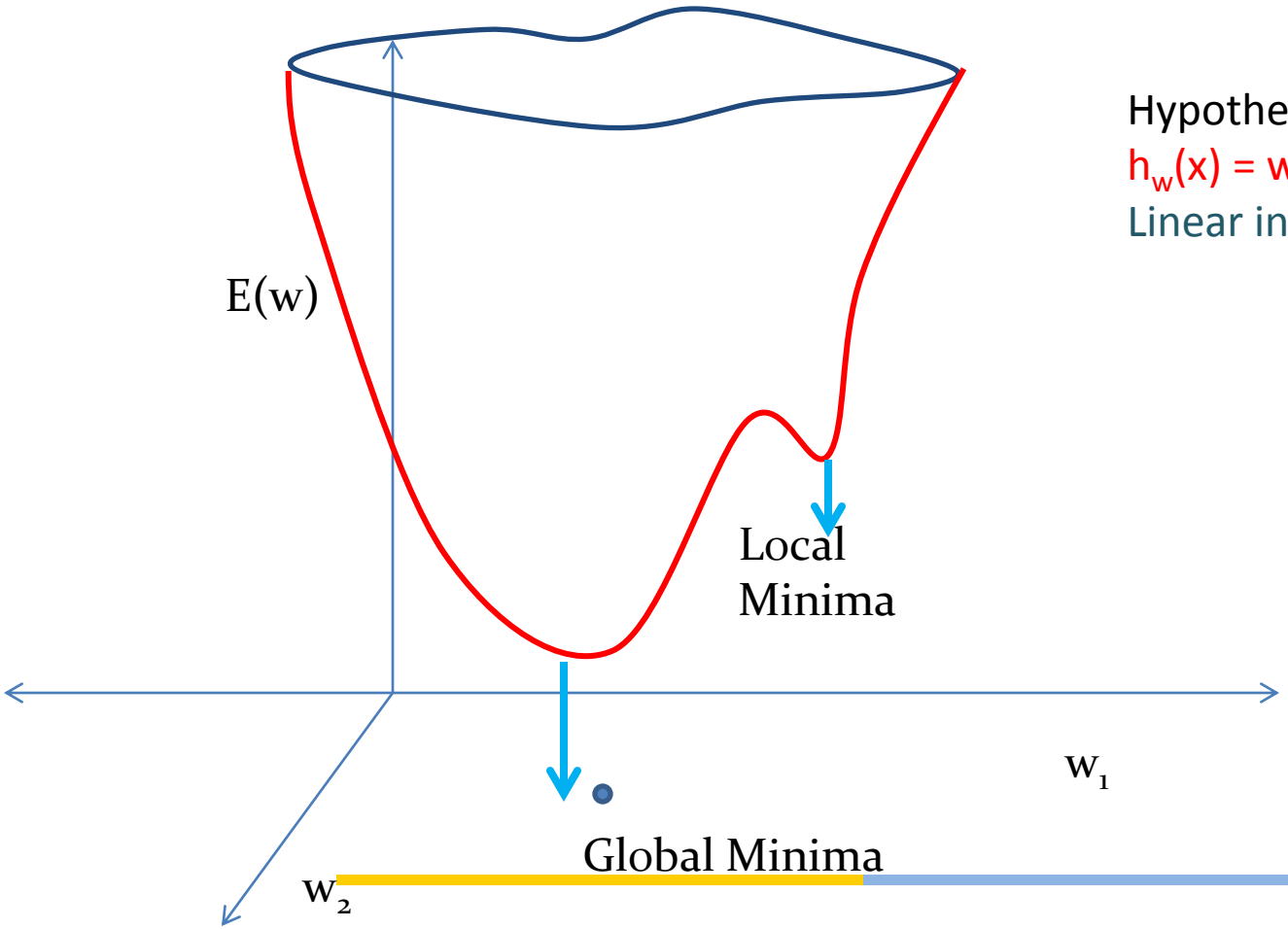
$h_w(x) = w_1x_1 + w_2x_2$ .....(1)

Linear in two variables

NOTE: each pair < w_1 ,w_2> corresponds to a line given by eq (1) while only one such pair corresponds to the line that best approximates the given training data

Another possible surface

Plotting error when $y=f(x_1,x_2)$

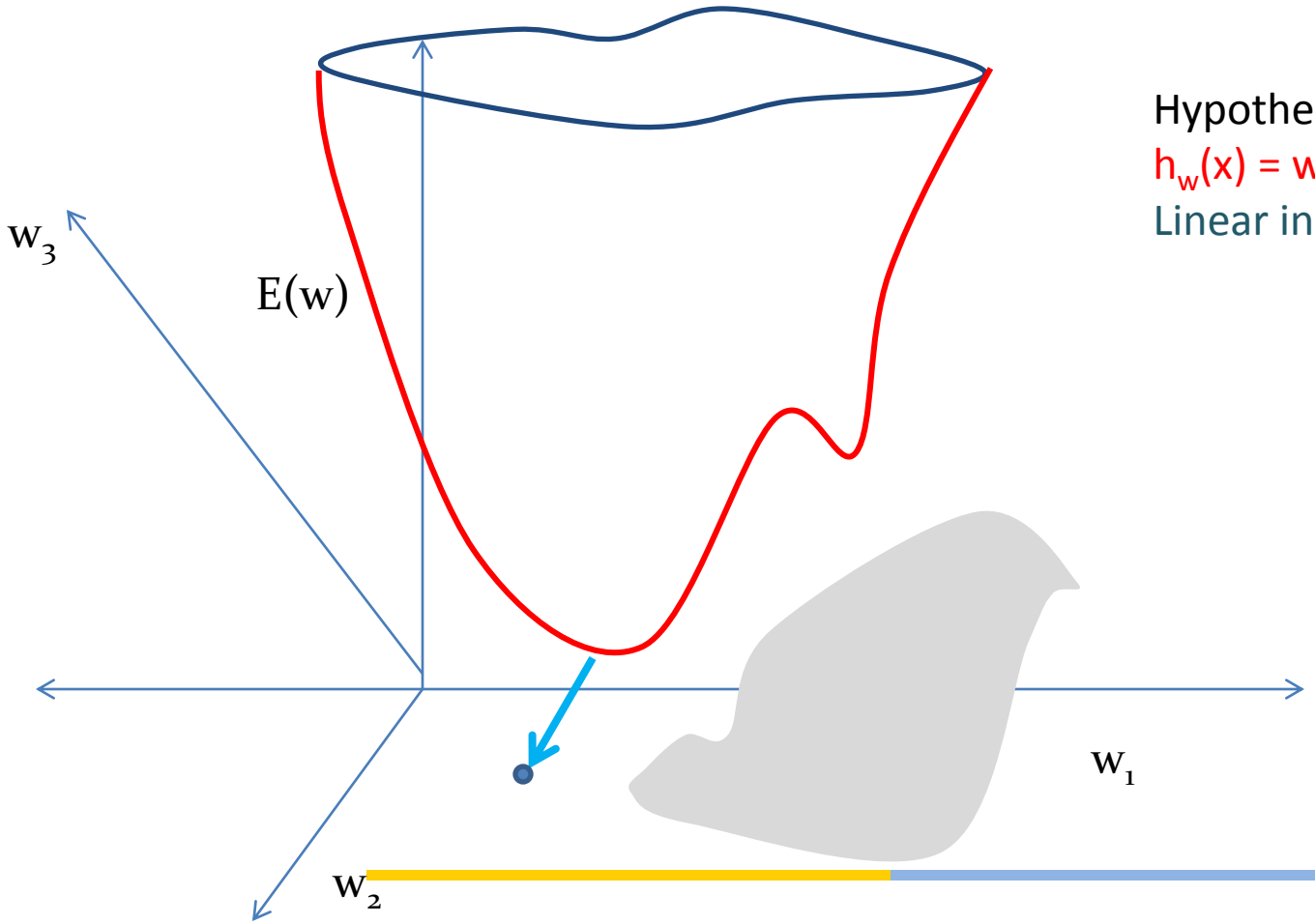


Hypothesis function

$h_w(x) = w_1x_1 + w_2x_2$ .....(1)

Linear in two variables

Difficult to visualize when $y=f(x_1,x_2, x_3)$



Hypothesis function

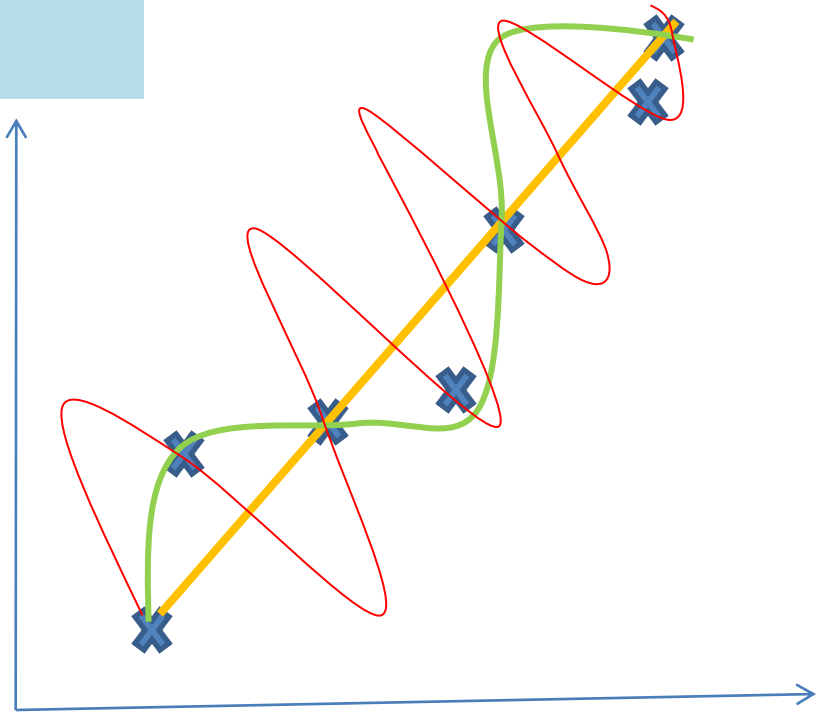
$h_w(x) = w_1x_1 + w_2x_2 + w_3x_3$

Linear in three variables

Learning of a function from given sample data- polynomial curve Learning

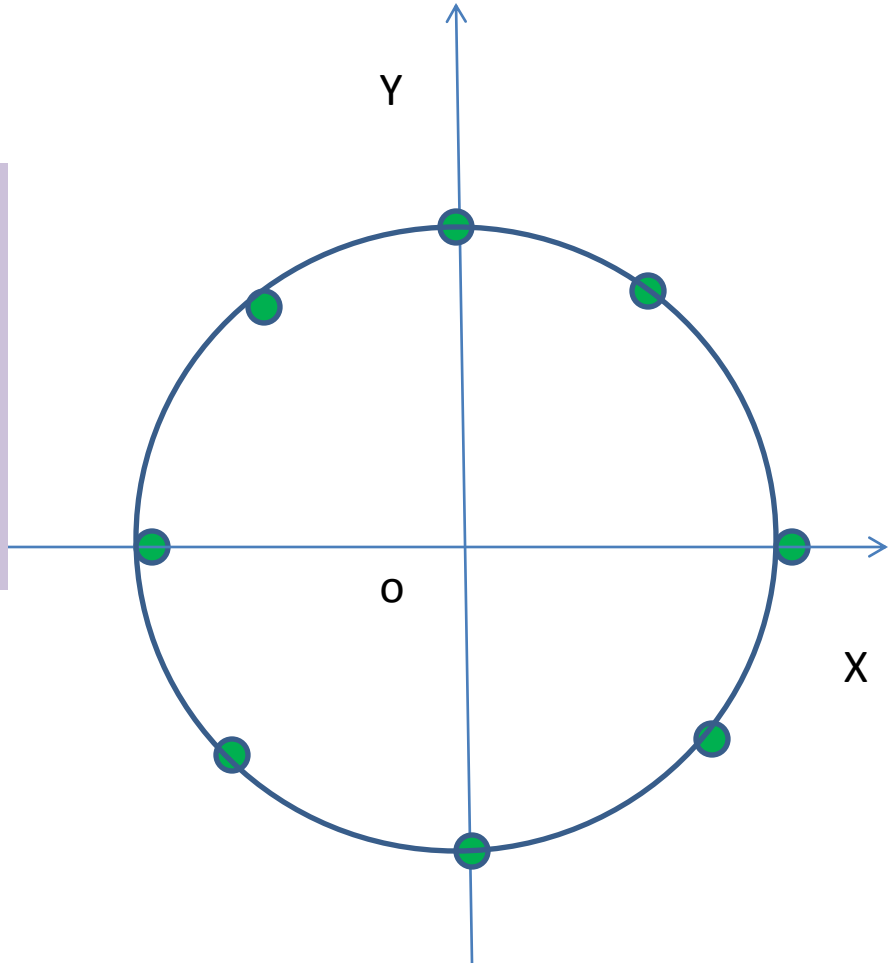
T: prediction of y-value for given x-value
P: least error
E: experience by training

1. Straight Line
2. Sinusoidal Curve
3. Other higher order polynomial



Generalization in Function Approximation

Generalization
If the NN answers -
-
What is f(-0.25)?
Or
f(0.001)
correctly

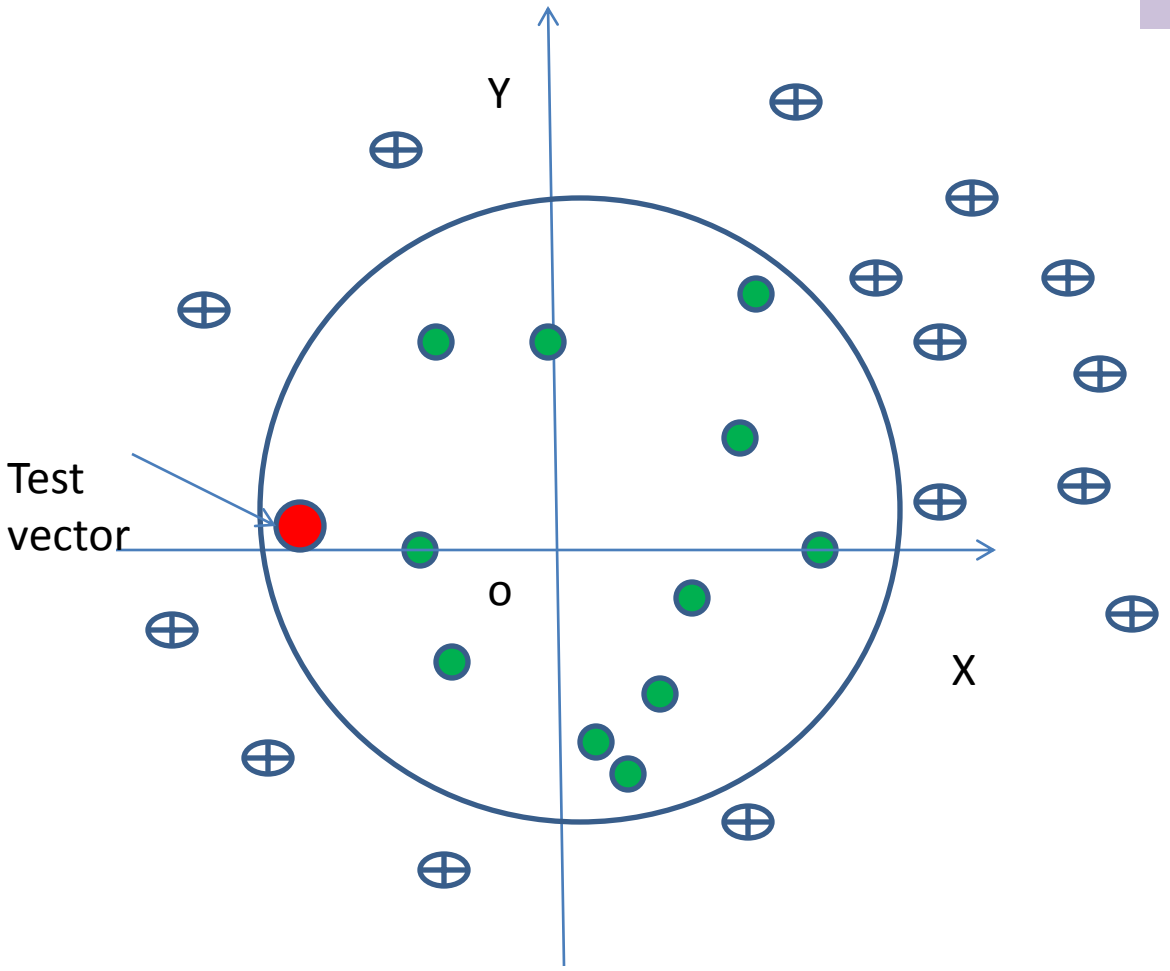


X	Y
1	0
0	1
0	-1
0.6	0.8
0.6	-0.8
-0.6	0.8
-0.6	-0.8
-1	0

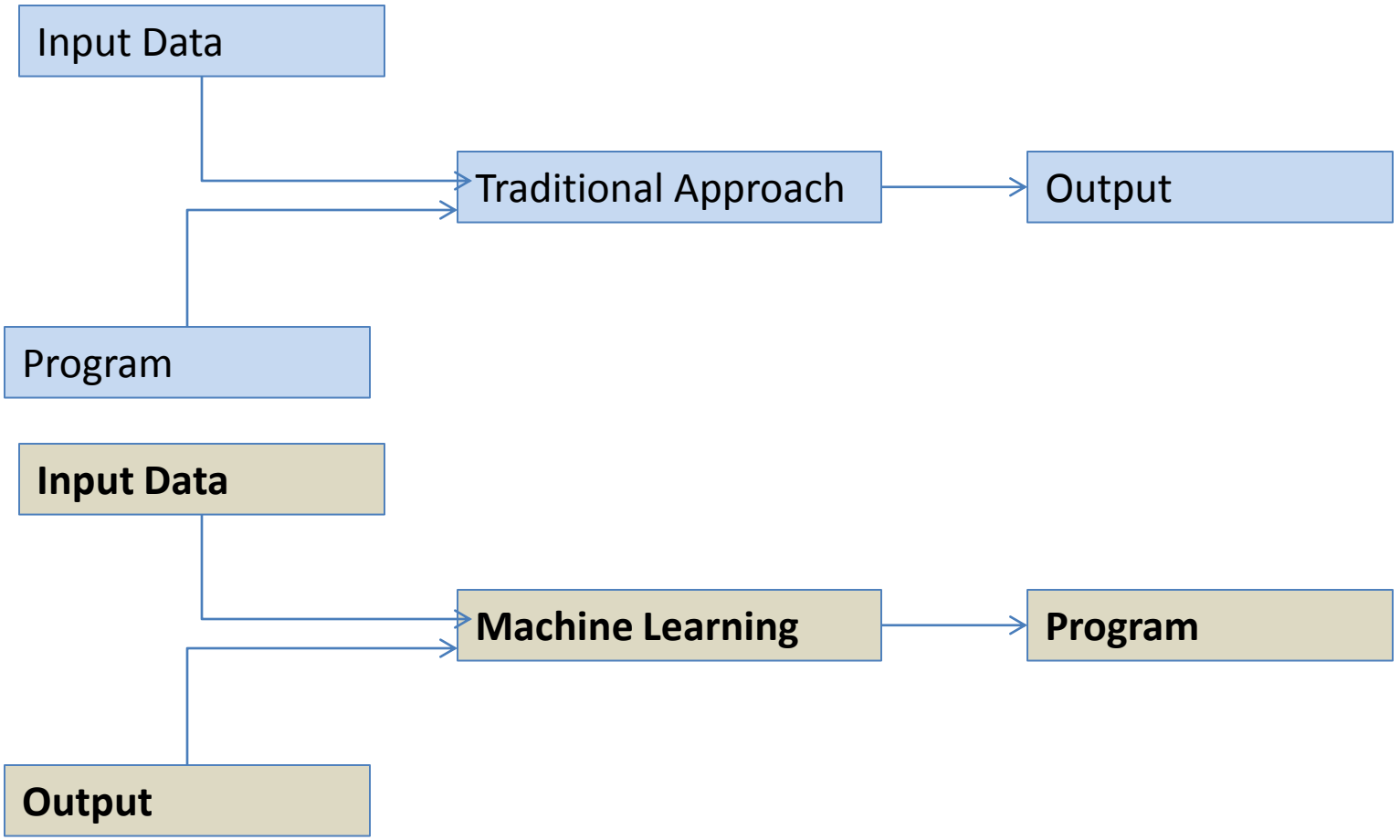
$Y = \pm \sqrt{(1-X^2)}$

Generalization in Classification Problem

Generalization
If the test feature vector can be correctly classified



Traditional Vs. Machine Learning



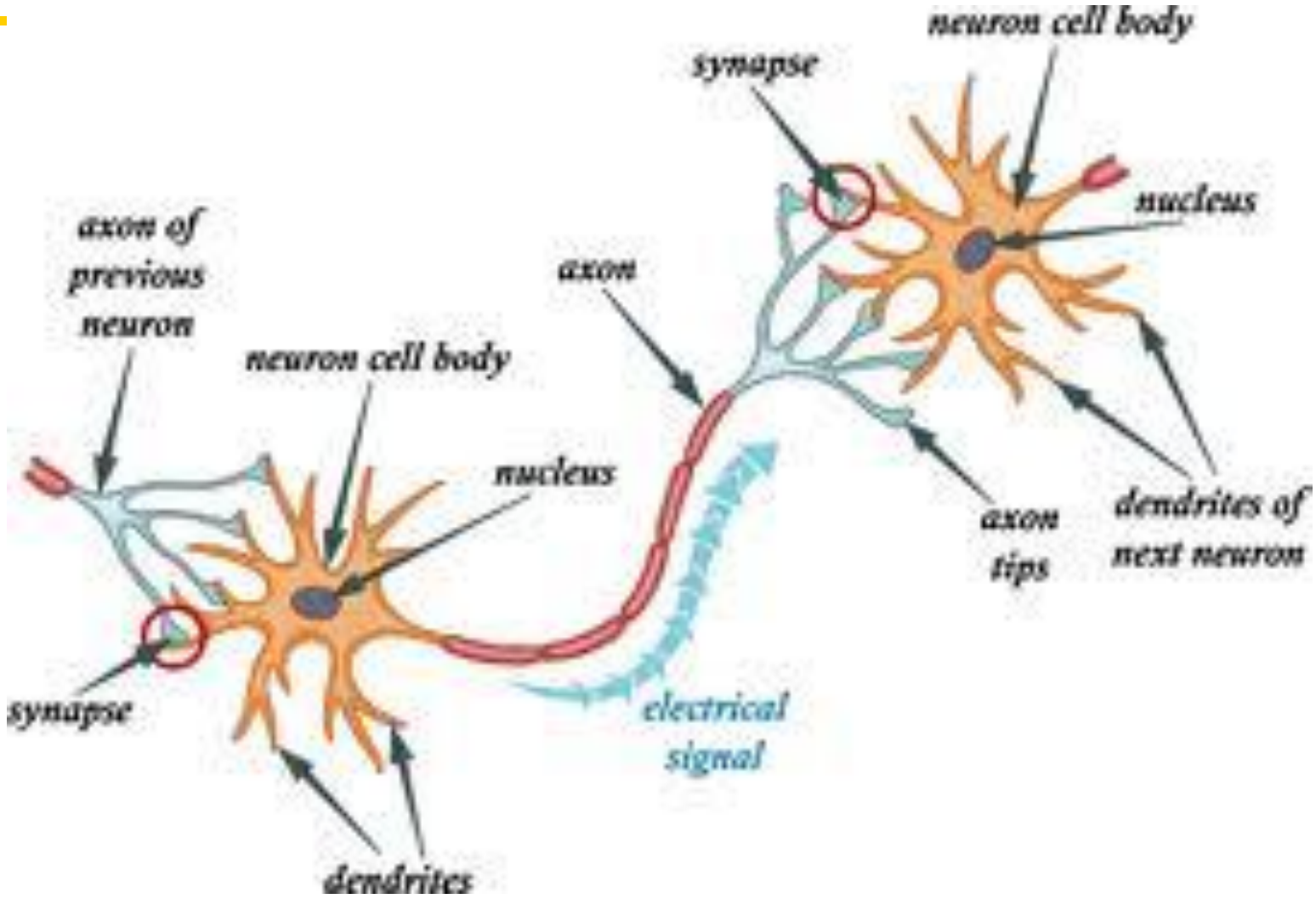
How does a program as an output realized?

- Program is characterized by its parameters.
- For example:
  - A neural network classifier is represented by its weights
  - Weights are obtained by analyzing input and output data
  - A decision tree is characterized by its attributes obtained by training input and output classes

Neural Networks

- Mathematical Models representing the massively parallel machines
- Model inspired by the working of human nervous system
- Has a number of neurons performing the task similar to human neuron
- Each neuron triggers the received input according to the weight.
- A neural network captures the environment it has to learn in terms of the weights.

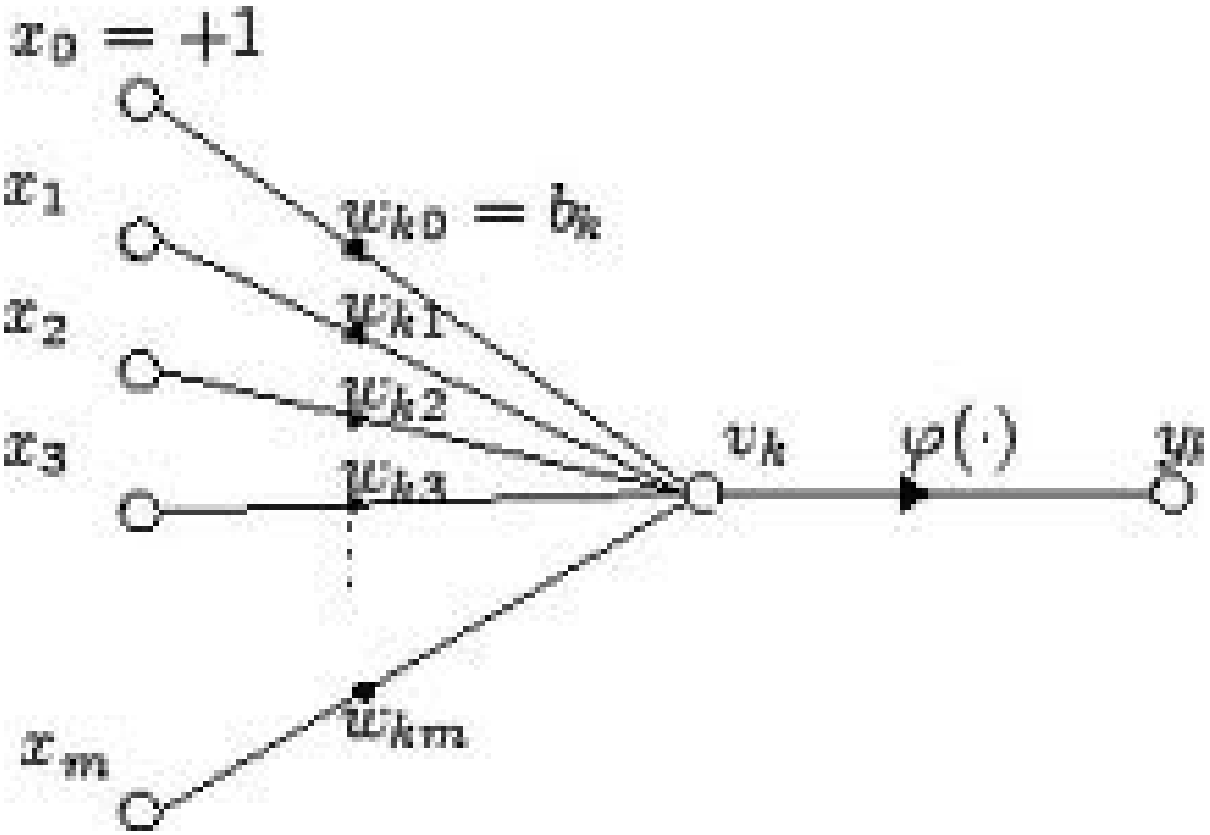
Biological Neuron



A Neuron

- A mathematical neuron is a processing unit capable of receiving inputs from single or multiple neurons and triggers a desired response.
- Each neuron has an associated activation function which takes as input the weighted sum of the inputs coming to the neuron and triggers a response depending on the associated threshold

A Mathematical Neuron



Artificial Neuron

- Weighted sum (n) = $\Sigma\ w_i x_i$
- Activation function f(n) = f($\Sigma\ w_i x_i$)
- Threshold T such that a response is triggered if f(n) > T