

Machine Learning:

Machine learning is a field of artificial intelligence that teaches computers to learn from data without being explicitly programmed.

Types of Machine learning :

1. Supervised Learning
2. Unsupervised Learning
3. Semi-Supervised Learning
4. Reinforcement Learning

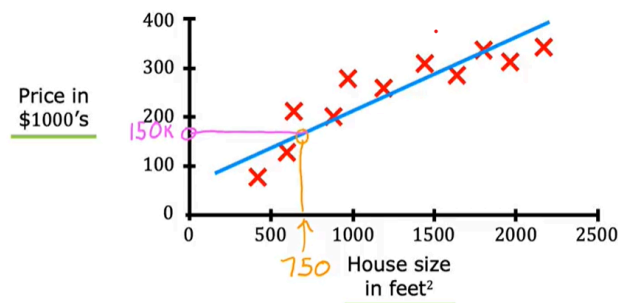
Supervised Learning :

In supervised learning, the algorithm learns from labeled data, meaning each training example has an input and a known correct output (label). The model makes predictions and adjusts based on the accuracy of those predictions.

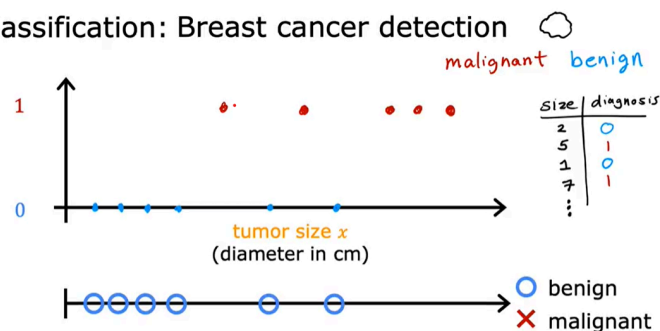
Types of Supervised Learning Tasks:

Task Type	Output Type	Example
Regression	Continuous Value	Predicting house prices, stock value
Classification	Discrete Class Label	Spam detection

Regression: Housing price prediction



Classification: Breast cancer detection

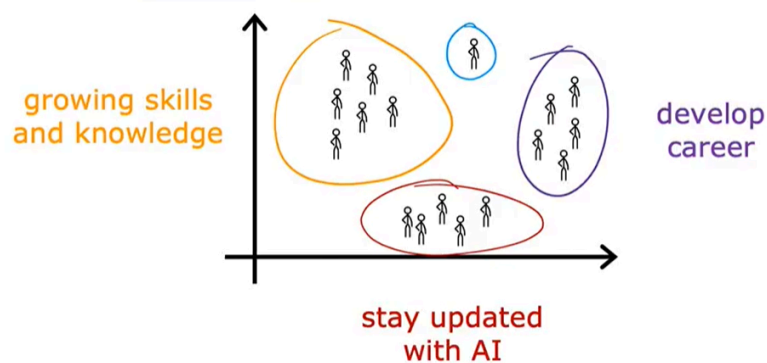


Unsupervised Learning :

Unsupervised learning deals with unlabeled data; the model explores data structure or patterns without predefined labels.

Main applications: Clustering, dimensionality reduction, anomaly detection.

Clustering: Grouping customers



Linear Regression Model:

Linear Regression is a supervised learning algorithm used to predict a continuous output variable based on one or more input features. It models the relationship between the input variables (X) and the output variable (y) using a straight line.

Terminology

Training Data used to train the model

set: x y
→ size in feet² → price in \$1000's

(1)	2104	400
(2)	1416	232
(3)	1534	315
(4)	852	178
...
(47)	3210	870

$x^{(1)} = 2104$ $y^{(1)} = 400$
 $(x^{(1)}, y^{(1)}) = (2104, 400)$
 $x^{(2)}$ $x^{(2)} \neq x^2$ not exponent

Notation:

x = "input" variable
feature

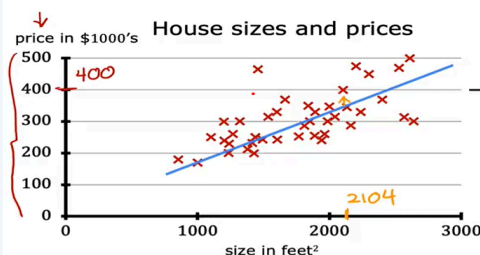
y = "output" variable
"target" variable

m = number of training examples

(x, y) = single training example

$(x^{(i)}, y^{(i)})$

$(x^{(i)}, y^{(i)})$ = i^{th} training example
(1st, 2nd, 3rd ...)

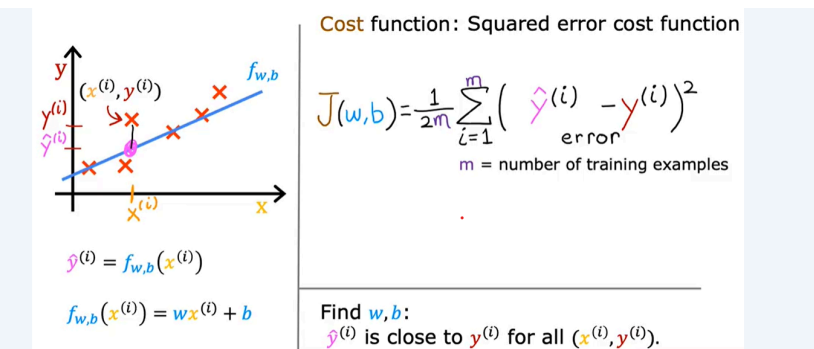


size in feet ²	price in \$1000's
2104	400
1416	232
1534	315
852	178
...	...
3210	870

Cost Function:

It calculates the difference between predicted values and actual values from the dataset.

 **Goal:** Minimize the cost function to improve model accuracy.



model:

$$f_{w,b}(x) = wx + b$$

parameters:

$$w, b$$

cost function:

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})^2$$

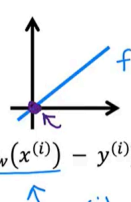
goal:

$$\text{minimize}_{w,b} J(w,b)$$

simplified

$$f_w(x) = wx$$

$$b = 0$$



$$J(w) = \frac{1}{2m} \sum_{i=1}^m (f_w(x^{(i)}) - y^{(i)})^2$$

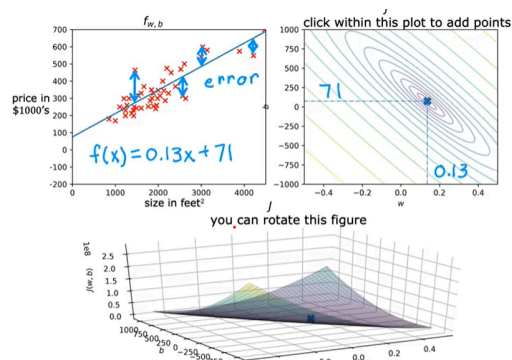
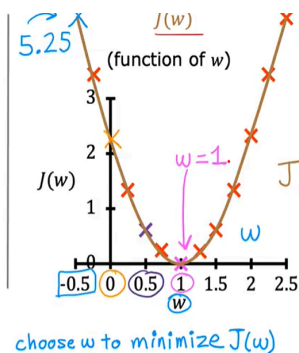
minimize $J(w)$

goal of linear regression:

$$\text{minimize}_w J(w)$$

general case:

$$\text{minimize}_{w,b} J(w,b)$$



Gradient Descent:

Gradient Descent is an optimization algorithm used to minimize the cost function (or loss function) in machine learning models, typically through iterative updates to model parameters.

repeat until convergence {

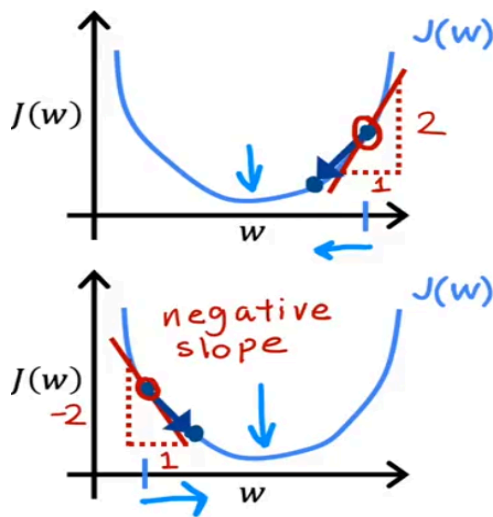
learning rate α derivative $\frac{\partial}{\partial w} J(w, b)$

$$\begin{cases} \underline{w} = w - \alpha \frac{\partial}{\partial w} J(w, b) \\ \underline{b} = b - \alpha \frac{\partial}{\partial b} J(w, b) \end{cases}$$

$$J(w)$$

$$w = w - \alpha \frac{\partial}{\partial w} J(w)$$

$$\min_w J(w)$$



$$w = w - \alpha \frac{d}{dw} J(w)$$

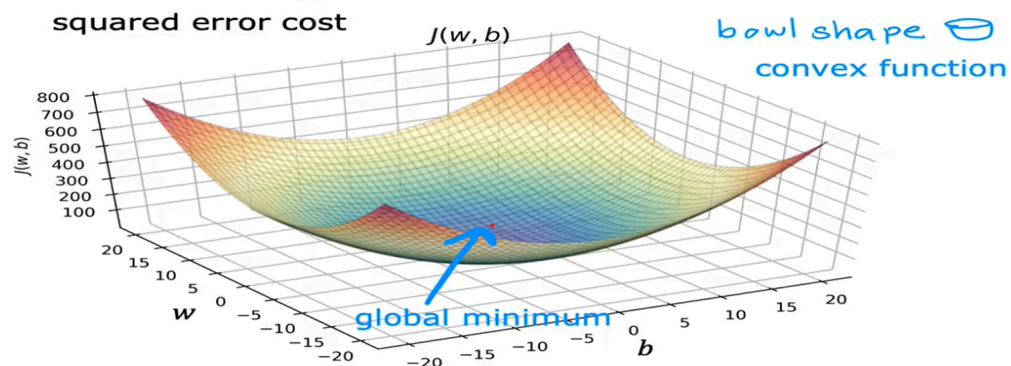
> 0

$$w = w - \alpha \cdot (\text{positive number})$$

$$\frac{d}{dw} J(w) < 0$$

$$w = w - \alpha \cdot (\text{negative number})$$

Gradient descent for linear regression



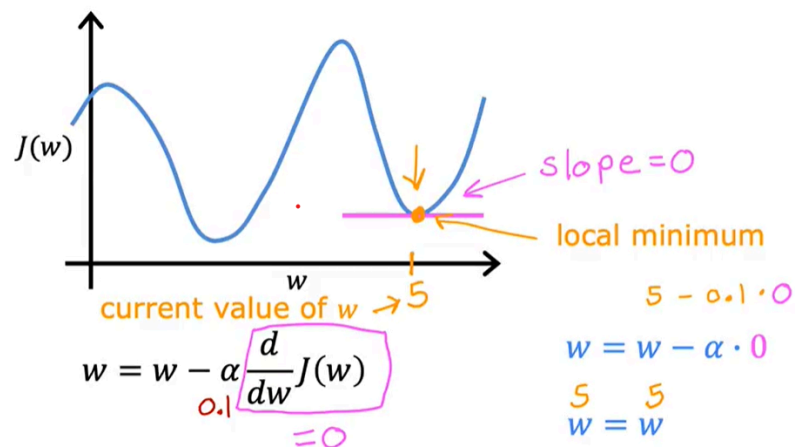
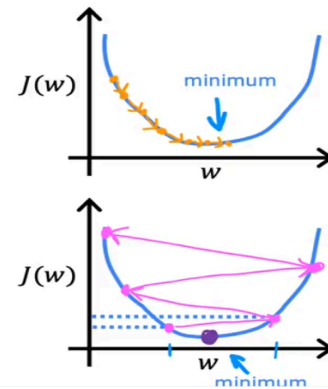
Learning Rate:

The learning rate, often denoted by α (alpha), is a hyperparameter that determines the size of the steps taken during gradient descent when updating model parameters.

$$w = w - \alpha \frac{d}{dw} J(w)$$

If α is too small...
Gradient descent may be slow.

If α is too large...
Gradient descent may:
- Overshoot, never reach minimum



Can reach local minimum with fixed learning rate α

$$w = w - \underbrace{\alpha \frac{d}{dw} J(w)}_{\text{smaller, not as large, large}}$$

Near a local minimum,
- Derivative becomes smaller
- Update steps become smaller

Can reach minimum without decreasing learning rate α

