Training Linear Regression

Gradient Decent

Function with more than one minimum

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

$$\frac{\partial J(w)}{\partial w} = 0$$

$$W_{new} = W_{old}$$

$\alpha \frac{\partial J(w)}{\partial w}$ larger => W_{change} large

Gradient decent algorithm

Linear regression model

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

Cost function

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

Gradient descent algorithm

repeat until convergence {

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

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

}

Derivative terms

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

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

Gradient decent algorithm

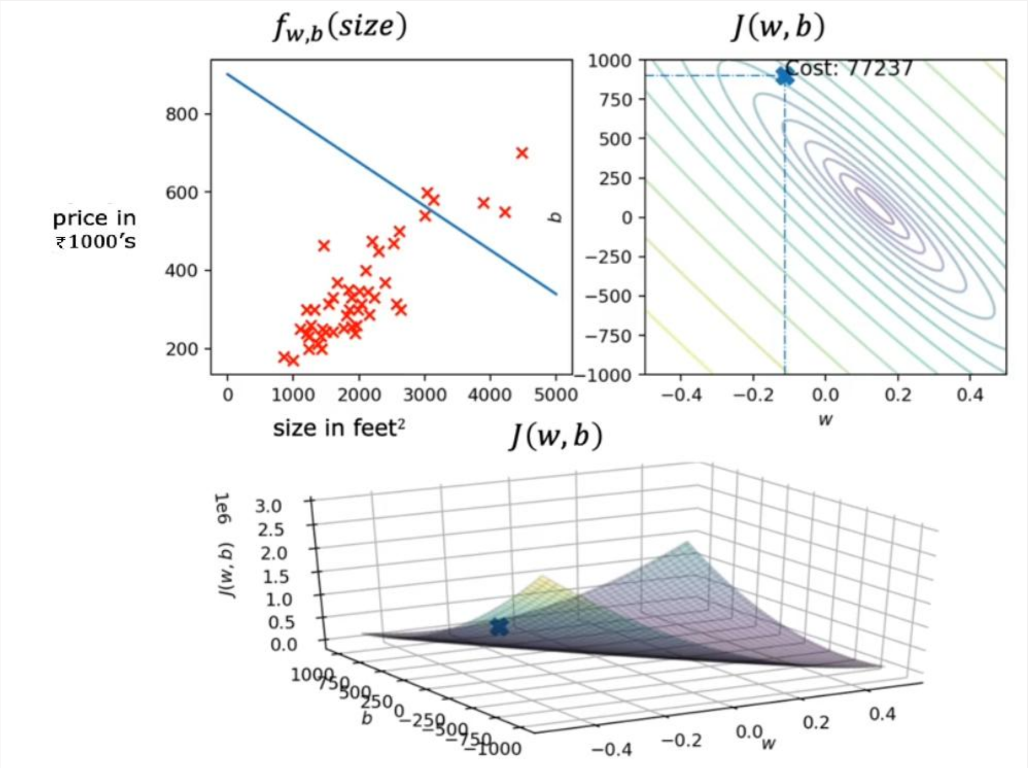
repeat until convergence {

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

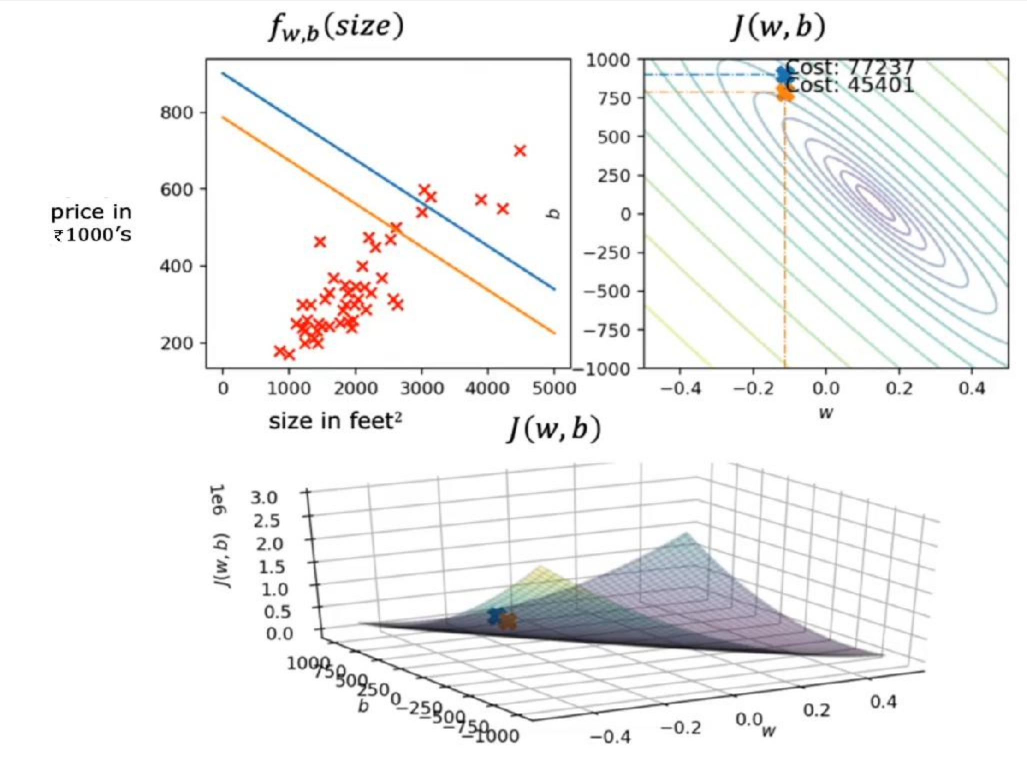
$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)})$$

}

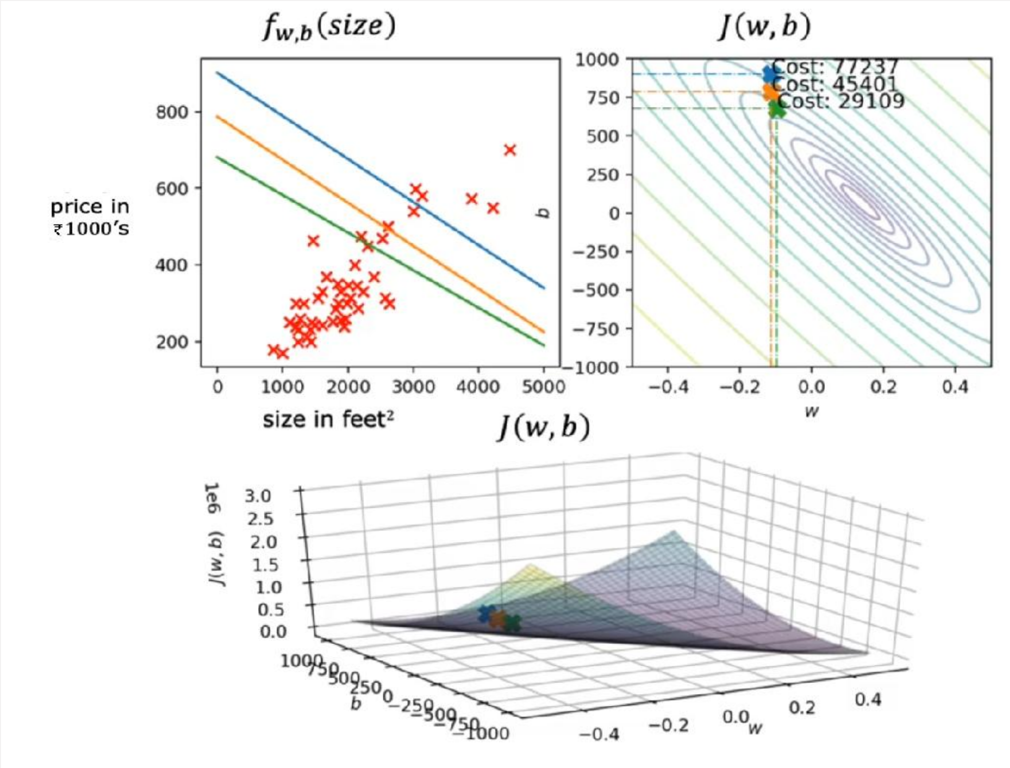
Gradient decent algorithm (in action)



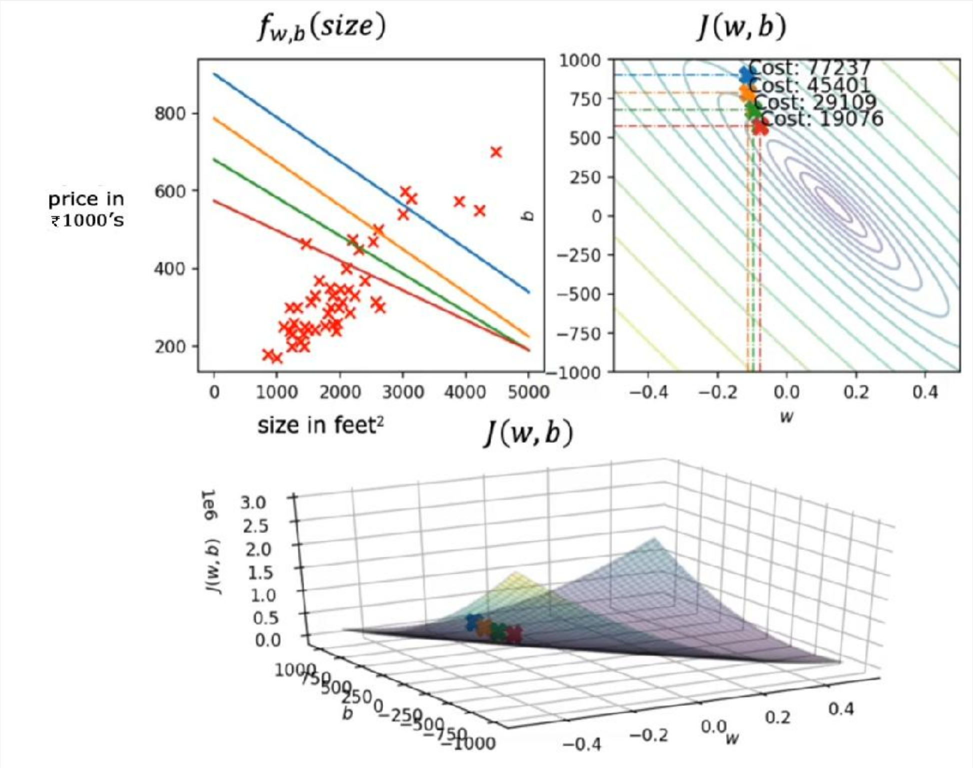
Gradient decent algorithm (in action)



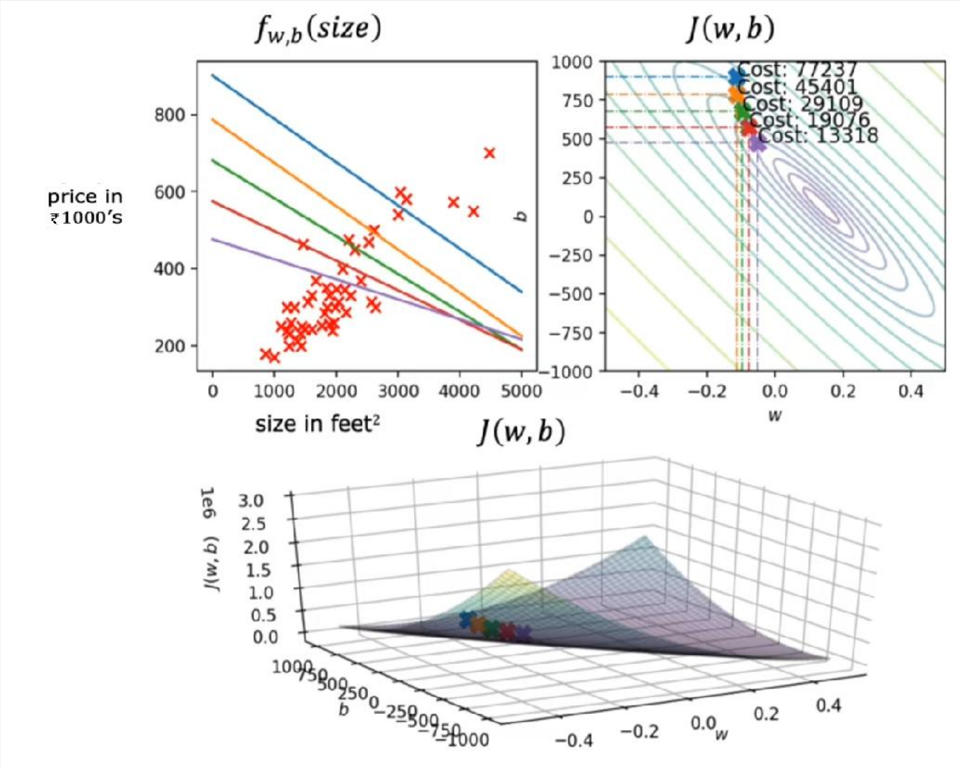
Gradient decent algorithm (in action)



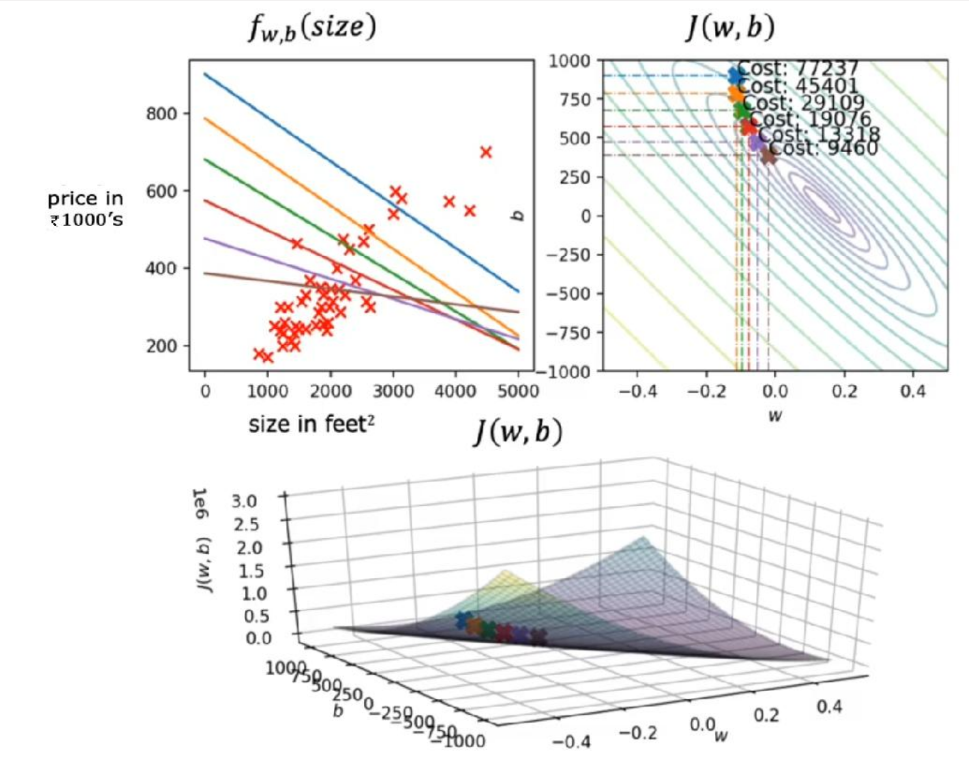
Gradient decent algorithm (in action)



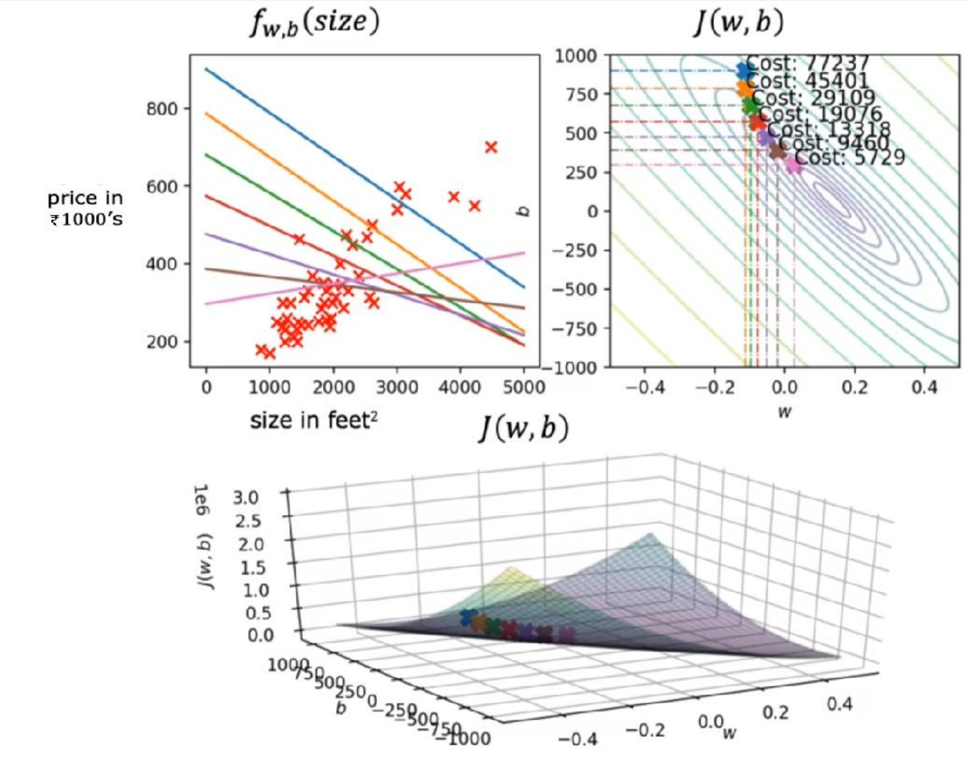
Gradient decent algorithm (in action)



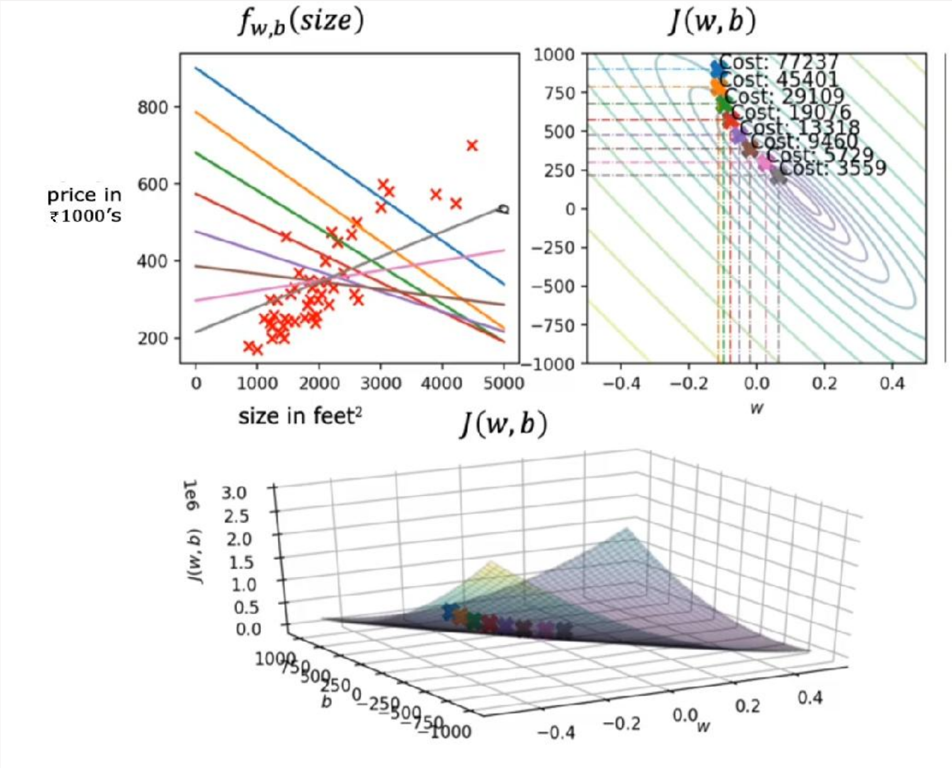
Gradient decent algorithm (in action)



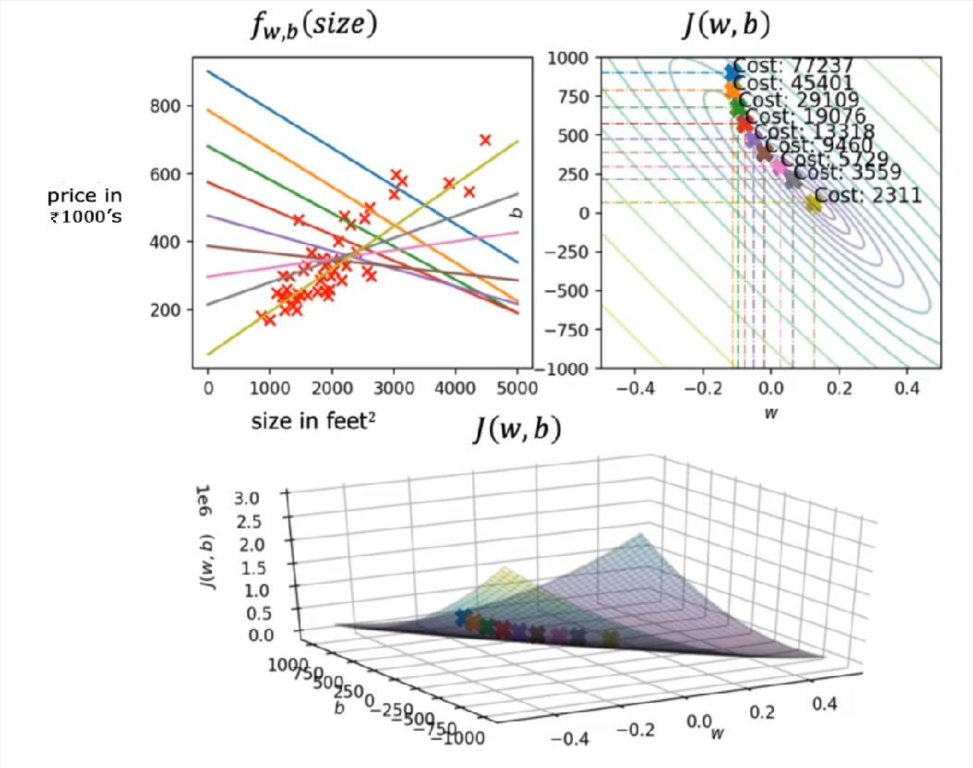
Gradient decent algorithm (in action)



Gradient decent algorithm (in action)



Gradient decent algorithm (in action)



Gradient descent

Batch

Mini-batch

x	y
1.0	2.0
2.0	4.1
3.0	6.0
4.0	8.1
5.0	10.0

Example

Dataset: Single Variable (x, y)

The relationship is approximately y=2x, but with slight noise to simulate real-world data.