Linear Regression

Intelligent Systems

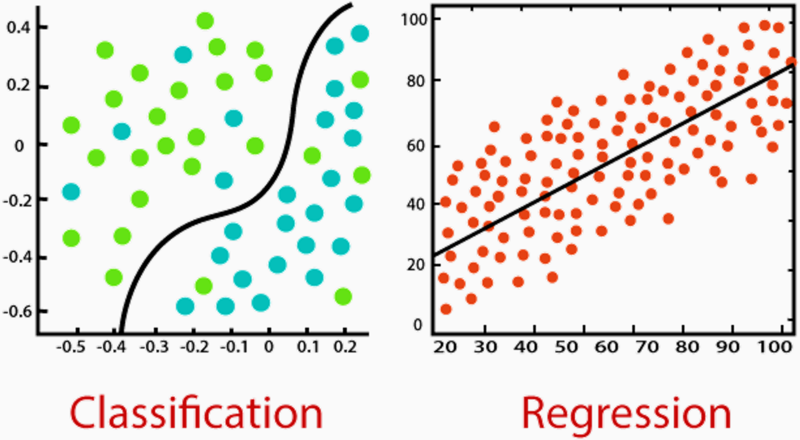
Rui Pedro Lopes

Based on the slides by:
Eduardo Bezerra (https://eic.cefet-rj.br/~ebezerra/gcc1932-am-2020-2/)

Outline

Introduction

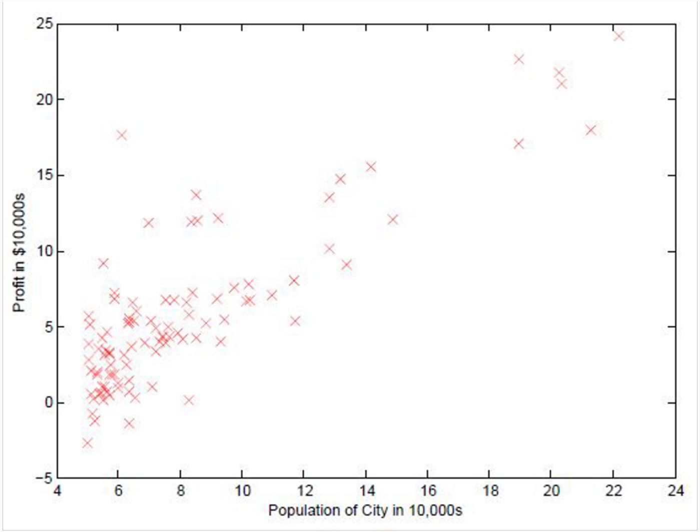
Linear Regression



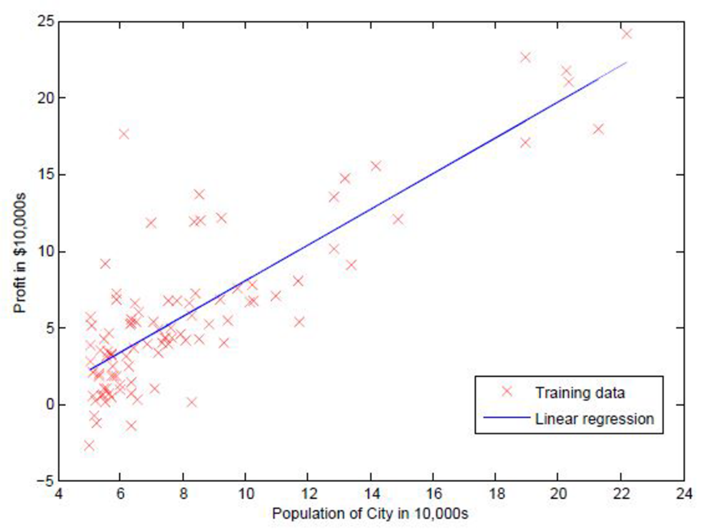
Linear Regression

The objective is to **induce** a regression model from a training set in which each example is represented by only one **feature** and the **target**

Linear Regression



Linear Regression

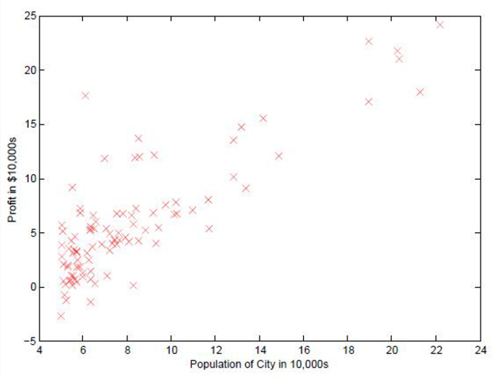


Representation of the hypothesis

Notation

- $m \rightarrow$ number of training examples
- $x^{(i)} \rightarrow$ feature value in the i^{th} training example
- $y^{(i)} \rightarrow$ target value in the i^{th} training example.

Representation of the hypothesis



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

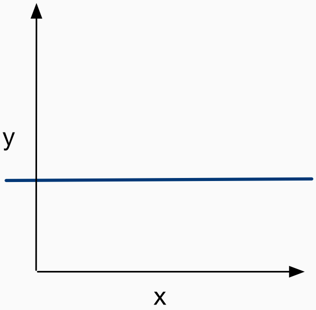
Cost Function

Model Parameters

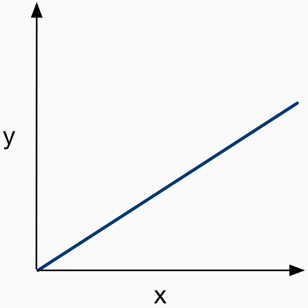
- Since we...
  - have a training set, and
  - defined the form (representation) of the hypothesis
- ... how do we determine the model **parameters**?

$$\theta_0, \theta_1$$

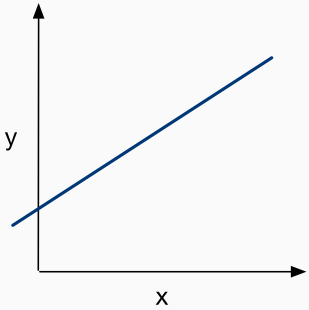
Model Parameters - Examples



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



$\theta_0=0$
$\theta_1=0.5$



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

Model Parameters

- How to determine the model parameters θ_0 and θ_1?
- Possibility: choose a combination of parameters that, for each $x^{(i)}$, the hypothesis produces a value close to the corresponding $y^{(i)}$
- But how to measure, objectively, if a parameter combination is better than others?

Cost Function

$$(h_\theta(x^{(i)}) - y^{(i)})$$

Cost Function

$$(h_\theta(x^{(i)}) - y^{(i)}) \rightarrow (h_\theta(x^{(i)}) - y^{(i)})^2$$

Cost Function

$$(h_\theta(x^{(i)}) - y^{(i)}) \rightarrow (h_\theta(x^{(i)}) - y^{(i)})^2 \rightarrow \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Cost Function

$$(h_\theta(x^{(i)}) - y^{(i)}) \rightarrow (h_\theta(x^{(i)}) - y^{(i)})^2 \rightarrow \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \rightarrow \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Cost Function

$$(h_\theta(x^{(i)}) - y^{(i)}) \rightarrow (h_\theta(x^{(i)}) - y^{(i)})^2 \rightarrow \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \rightarrow \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \text{ Mean Squared Error (MSE)}$$

Cost Function

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

Geometric Intuition

- Linear Regression: “finding parameter values that minimize the cost function“
- What is the geometric intuition?

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

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

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

Just one parameter

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

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

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

Just one parameter

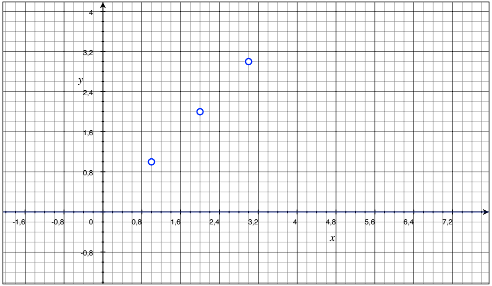
$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_1(x^{(i)}) - y^{(i)})^2 \tag{1}$$

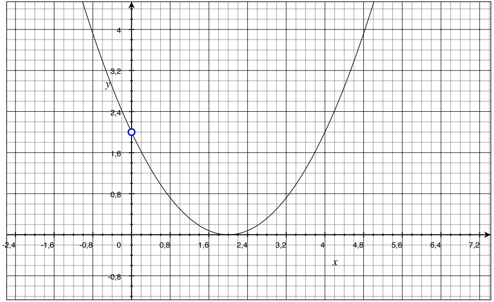
$$= \frac{1}{2m}\sum_{i=1}^{m}(\theta_1^2(x^{(i)})^2 - 2\theta_1 x^{(i)} y^{(i)} + (y^{(i)})^2) \tag{2}$$

$$= \left(\frac{1}{2m}\sum_{i=1}^{m}(x^{(i)})^2\right)\theta_1^2 + \left(-2\frac{1}{2m}\sum_{i=1}^{m}x^{(i)}y^{(i)}\right)\theta_1 + \left(\frac{1}{2m}\sum_{i=1}^{m}(y^{(i)})^2\right) \tag{3}$$

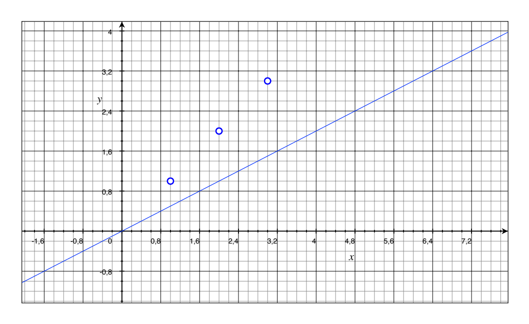
$$= a\theta_1^2 + b\theta_1 + c \tag{4}$$

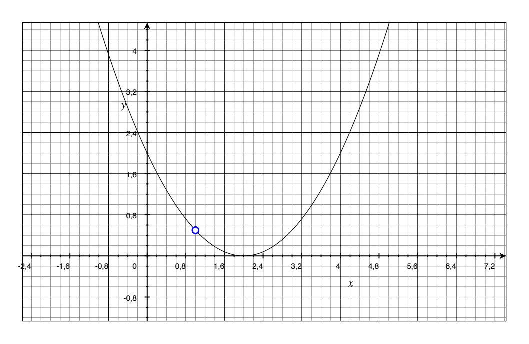
Just one parameter $\theta = 0$



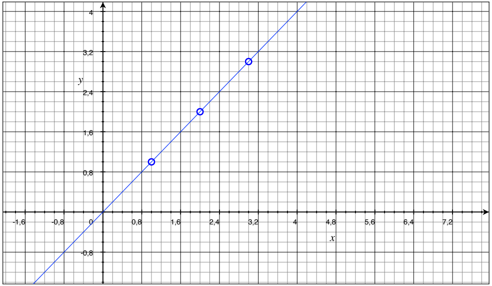


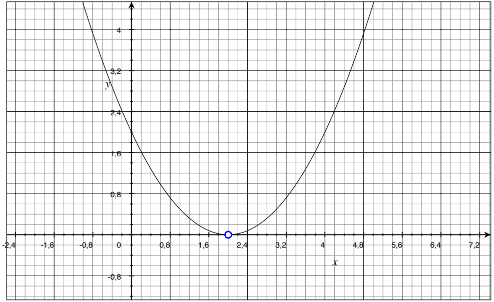
Just one parameter $\theta = 0.5$



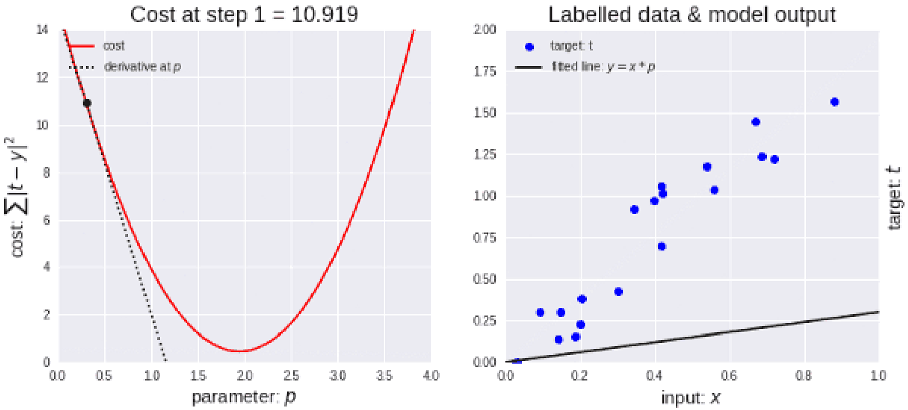


Just one parameter $\theta = 1$

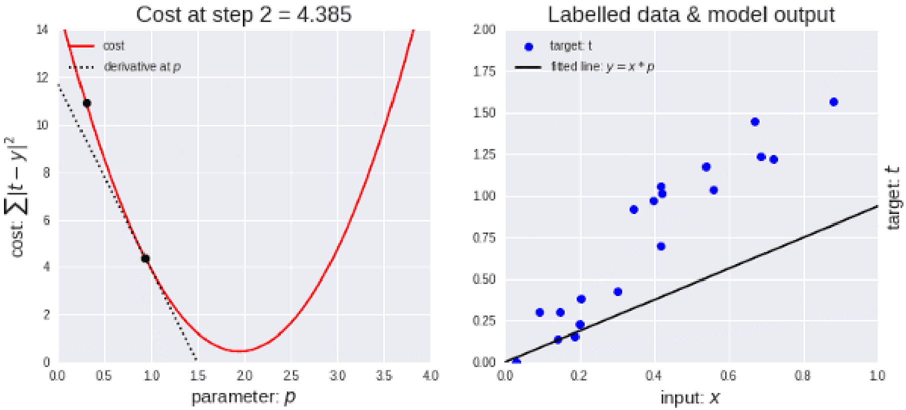




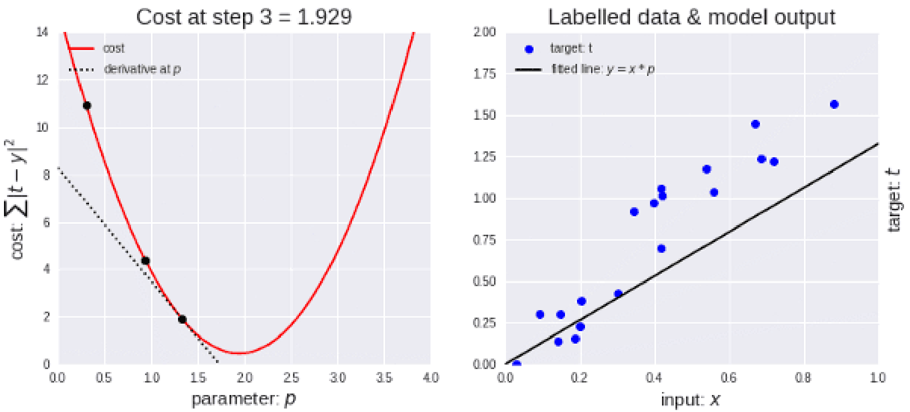
Geometric Intuition



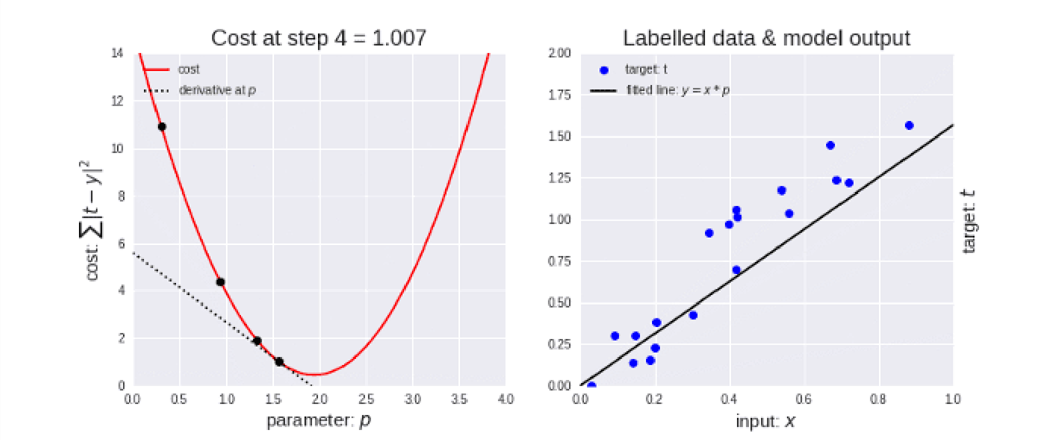
Geometric Intuition



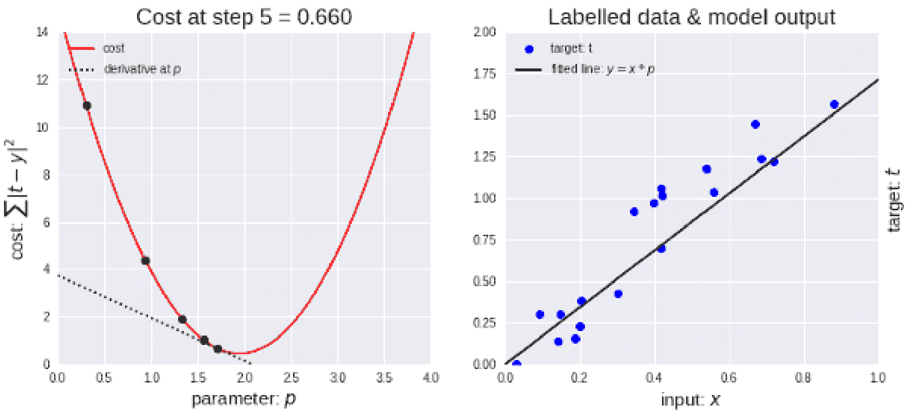
Geometric Intuition



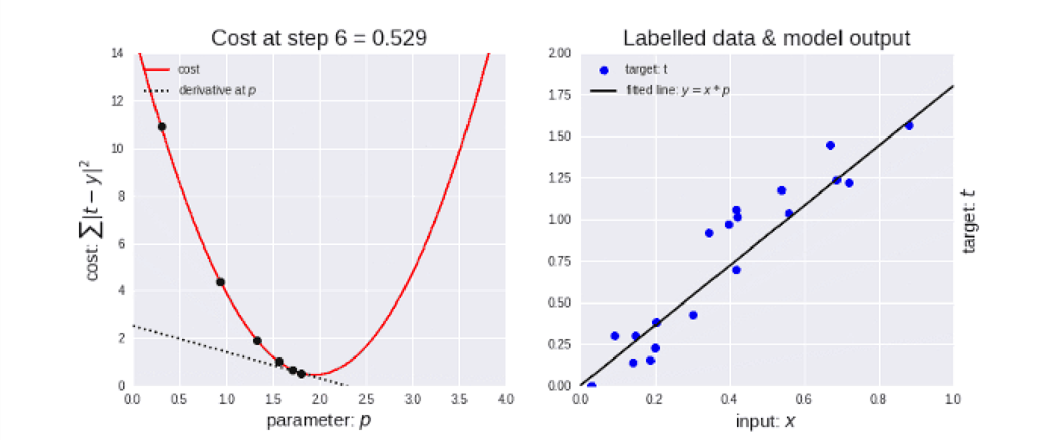
Geometric Intuition



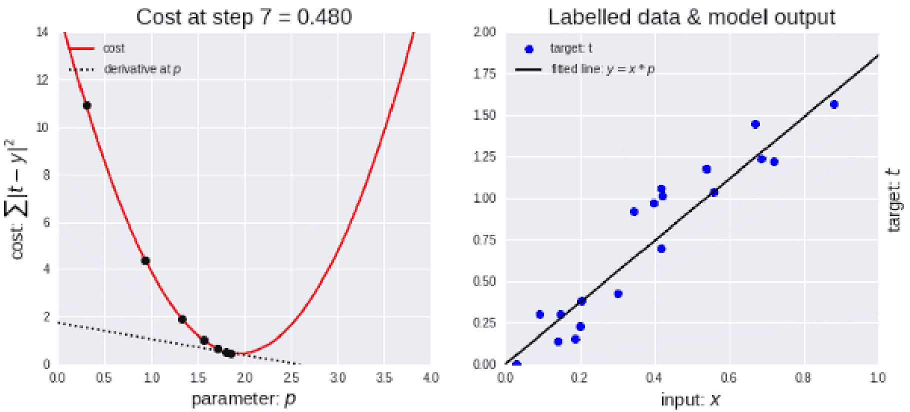
Geometric Intuition



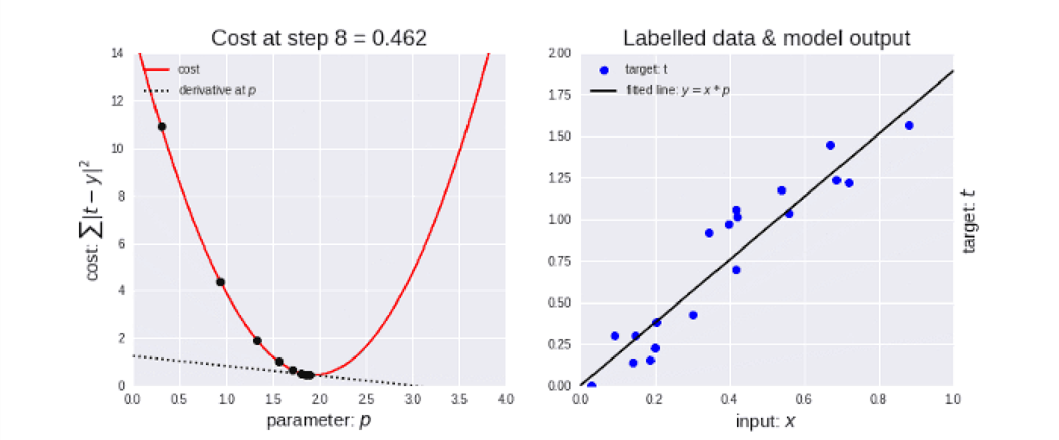
Geometric Intuition



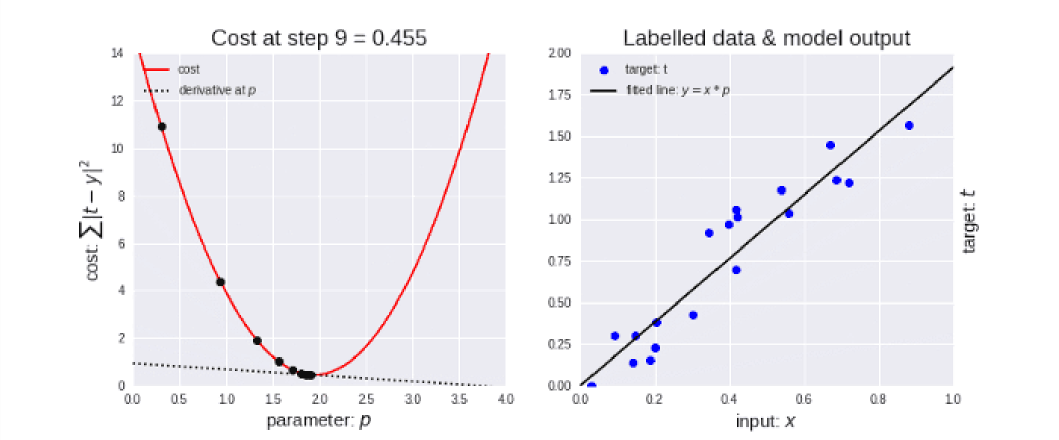
Geometric Intuition



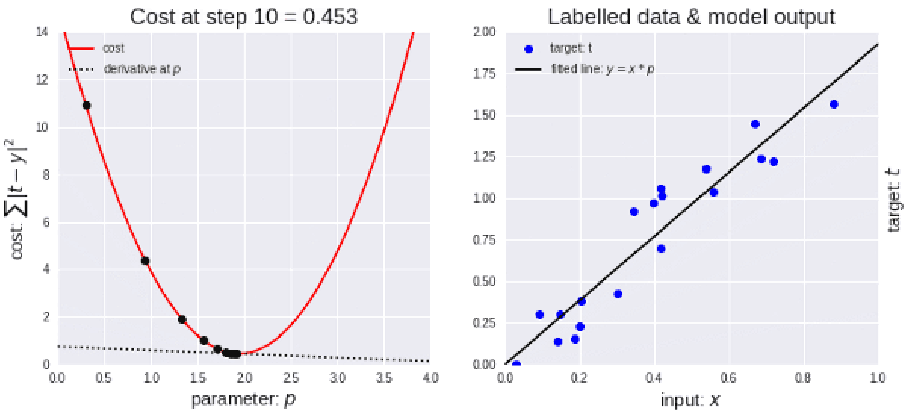
Geometric Intuition



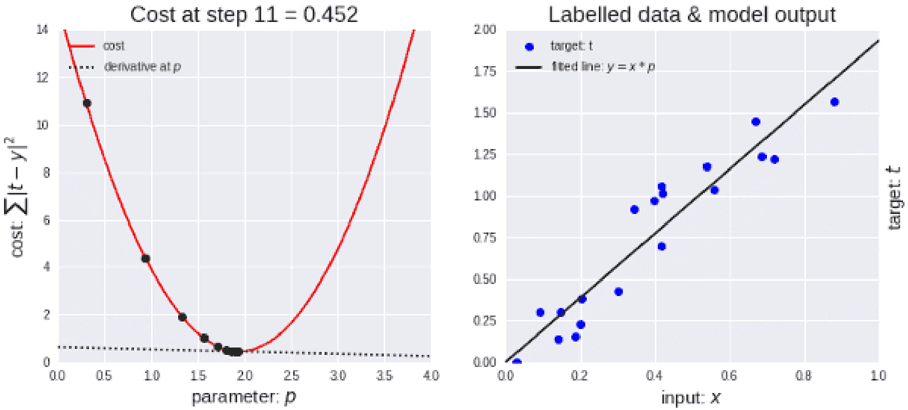
Geometric Intuition



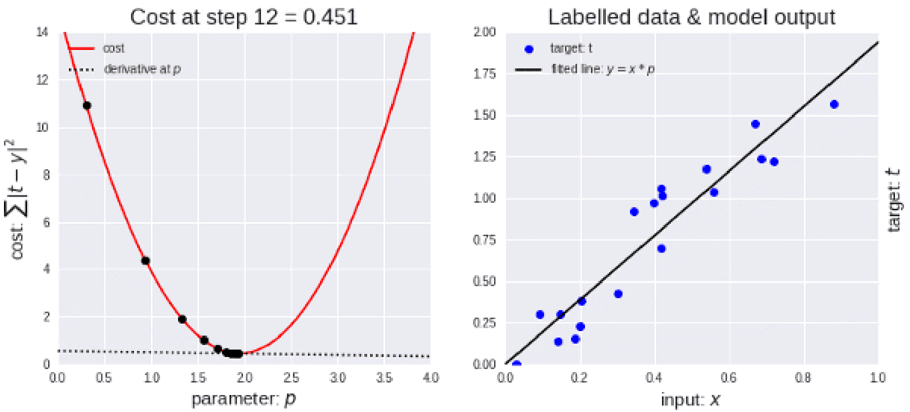
Geometric Intuition



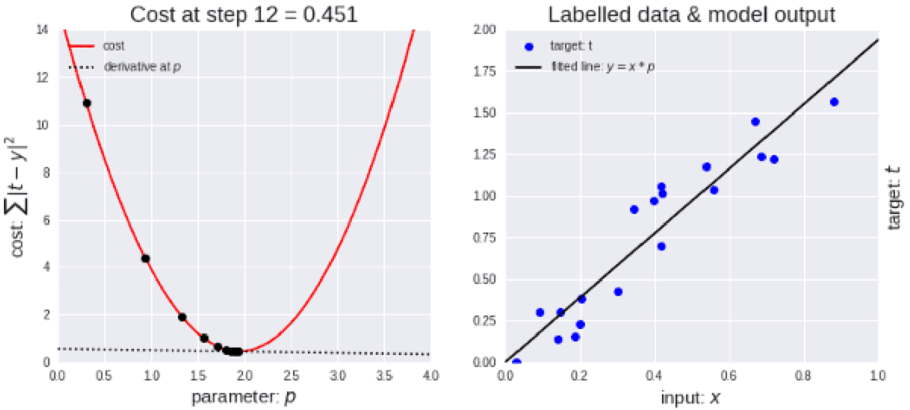
Geometric Intuition



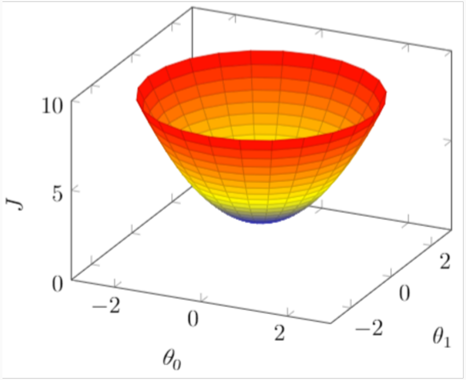
Geometric Intuition



Geometric Intuition

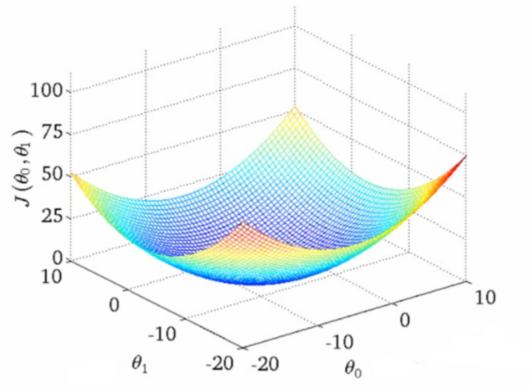


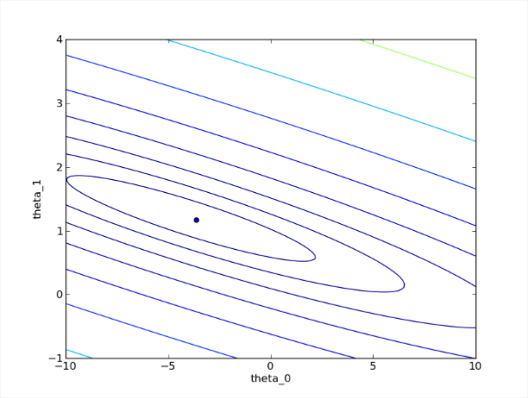
Two Parameters



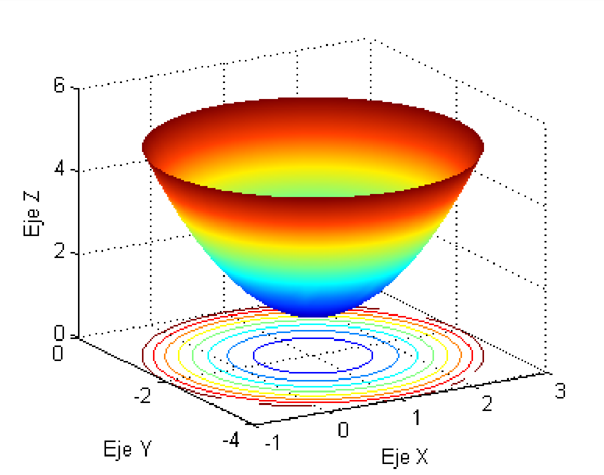
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Level curves of the function J (two parameters





Level curves of the function J (two parameters)



Parameter Learning

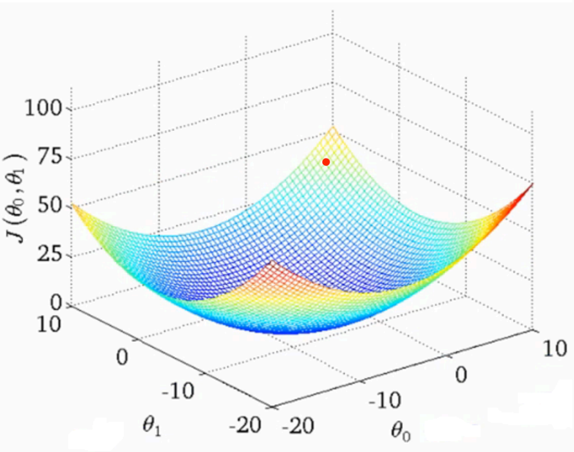
Gradient Descent

- Problem: Given a cost function J, determine the combination of parameter values that minimizes J

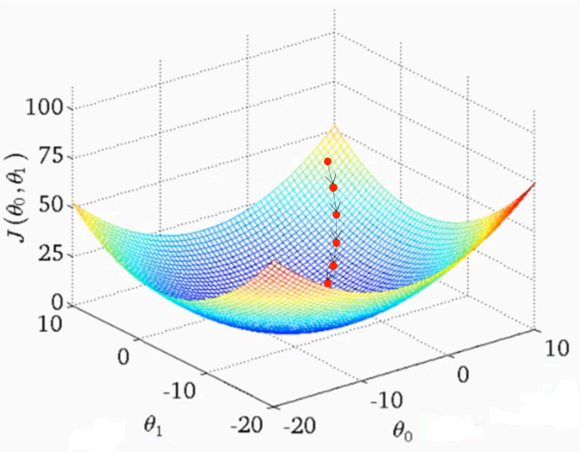
$$\min_{\theta_0,\theta_1,\ldots,\theta_n} J(\theta_0, \theta_1, \ldots, \theta_n)$$

- Algorithm (for linear regression with single variable, $n = 2$)
  1. Initialize the parameters θ_0 and θ_1
  2. Iteractively, change θ_0 and θ_1 with the purpose of finding the minimum value for $J(\theta_0, \theta_1)$

Initialize the parameters θ_0 and θ_1



Iteractivelly change the parameters θ_0 and θ_1



Gradient Descent

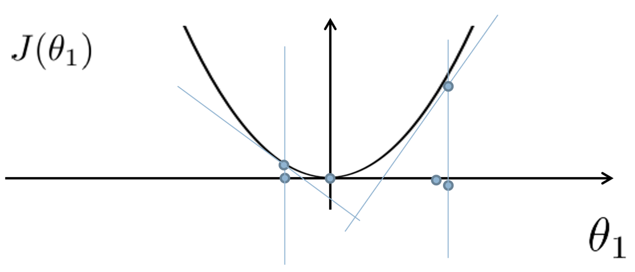
```
repeat until convergence: {
```

$$\theta_j \leftarrow \theta_j - \alpha \frac{\delta}{\delta \theta_j} J(\theta)$$

```
}
```

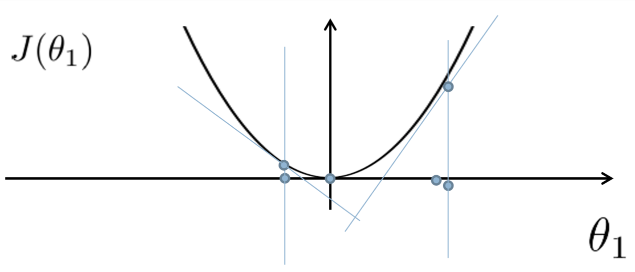
$$\text{temp0} := \theta_0 - \alpha \frac{\delta}{\delta \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\delta}{\delta \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

GD with one parameter



$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\delta}{\delta \theta_1} J(\theta_1)$$

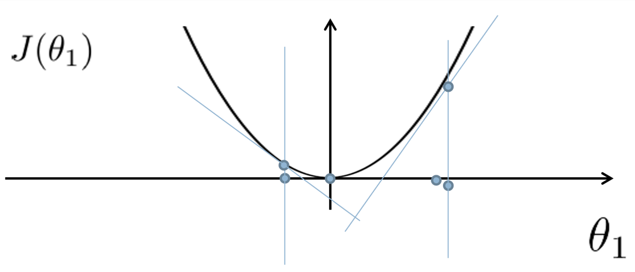
GD with one parameter



$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\delta}{\delta\theta_1} J(\theta_1)$$

1. case:
   $\frac{\delta}{\delta\theta_1} J(\theta_1) < 0 \Rightarrow \theta_1$ increases

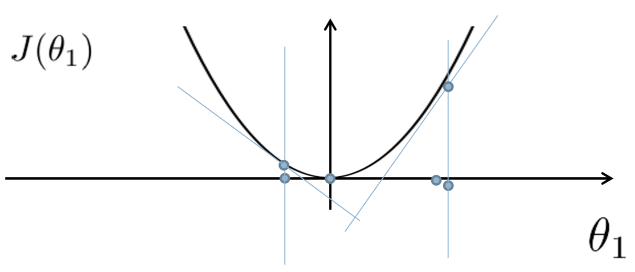
GD with one parameter



$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\delta}{\delta \theta_1} J(\theta_1)$$

1. case:
   $\frac{\delta}{\delta\theta_1} J(\theta_1) < 0 \Rightarrow \theta_1$ increases
2. case:
   $\frac{\delta}{\delta\theta_1} J(\theta_1) > 0 \Rightarrow \theta_1$ decreases

GD with one parameter

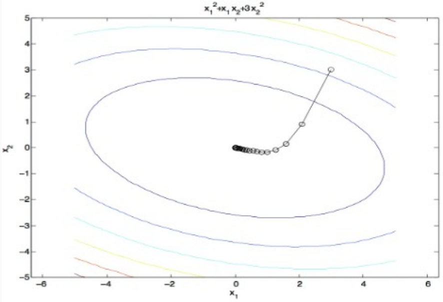


$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\delta}{\delta\theta_1} J(\theta_1)$$

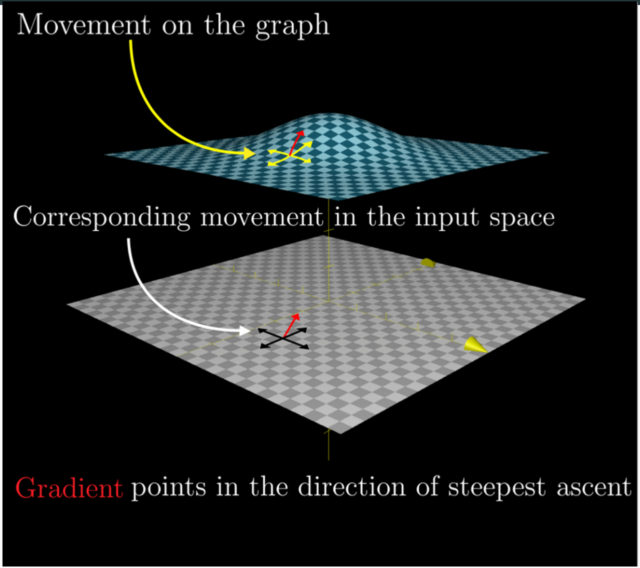
1. case:
   $\frac{\delta}{\delta\theta_1} J(\theta_1) < 0 \Rightarrow \theta_1$ increases
2. case:
   $\frac{\delta}{\delta\theta_1} J(\theta_1) > 0 \Rightarrow \theta_1$ decreases
3. case:
   $\frac{\delta}{\delta\theta_1} J(\theta_1) = 0 \Rightarrow \theta_1$ unchanged

GD with two parameters

- Idea:
  - Start anywhere
  - Repeat: take a step in the steepest descent direction



Steepest descent direction



$$\nabla J = \begin{bmatrix} \frac{\delta J(\theta)}{\delta \theta_0} \\ \frac{\delta J(\theta)}{\delta \theta_1} \end{bmatrix}$$

Steepest descent direction

- In the general case (i.e., n parameters), the steepest descent direction corresponds to the reverse direction of the cost function gradient

$$\nabla J = \begin{bmatrix} \frac{\delta J(\theta)}{\delta \theta_0} \\ \frac{\delta J(\theta)}{\delta \theta_1} \\ \vdots \\ \frac{\delta J(\theta)}{\delta \theta_n} \end{bmatrix}$$

GD for Linear Regression

GD for Linear Regression

Gradient Descent

repeat until convergence: {

$\theta_j = \theta_j - \alpha \frac{\delta}{\delta \theta_j} J(\theta_0, \theta_1)$

(for $j = 1$ and $j = 0$)

}

Linear Regression

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

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

GD for Linear Regression (one variable)

```
repeat until convergence: {
```

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

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

```
}
```

In practice...

Linear regression demo

Linear regression exercise