Loss Function

Regression: Output a Scalar

- Stock market forecast

$f($  $) =$ Dow Jones Industrial Average at tomorrow

- Self-driving car

$f($  $) =$ Steering wheel direction

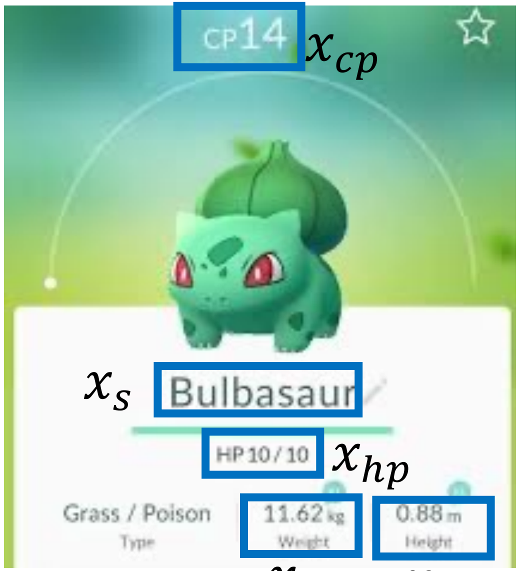
- Recommendation

$f($ Consumer A Product B $) =$ Purchasing likelihood

Example Application

Estimating the Combat Power (CP) of a Pokémon after evolution

$f($



x_{cp} x_s x_{hp} x_w x_h

x

$) =$ CP after evolution y

Step One: Model

$y = b + w \cdot x_{cp}$

w and b are parameters (can be any value)

A set of function

Model

$f_1, f_2 \cdots$

f_1: $y = 10.0 + 9.0 \cdot x_{cp}$

f_2: $y = 9.8 + 9.2 \cdot x_{cp}$

f_3: $y = -0.8 - 1.2 \cdot x_{cp}$

…… infinite

$f($ 

$x) =$ CP after evolution y

Linear model: $y = b + \sum w_i x_i$

x_i: $x_{cp}, x_{hp}, x_w, x_h \ldots$

Feature

w_i: weight, b: bias

Step Two: Goodness of Function

$y = b + w \cdot x_{cp}$

A set of function

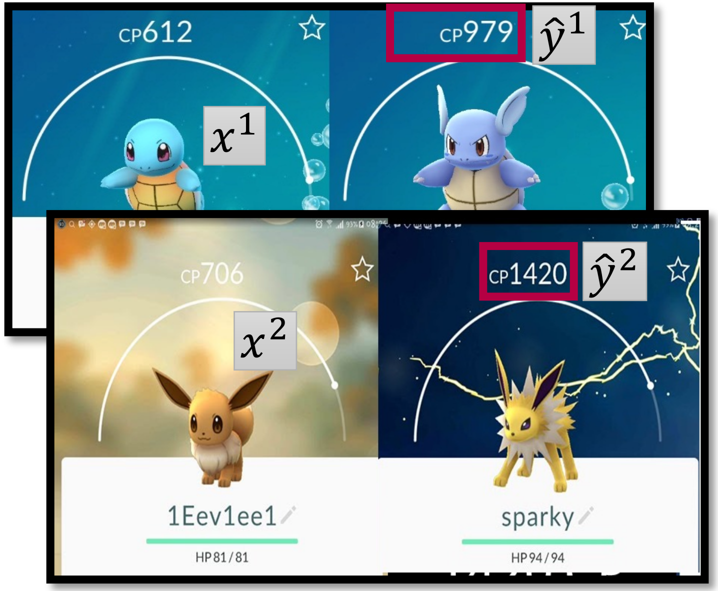
Model

$f_1, f_2 \cdots$

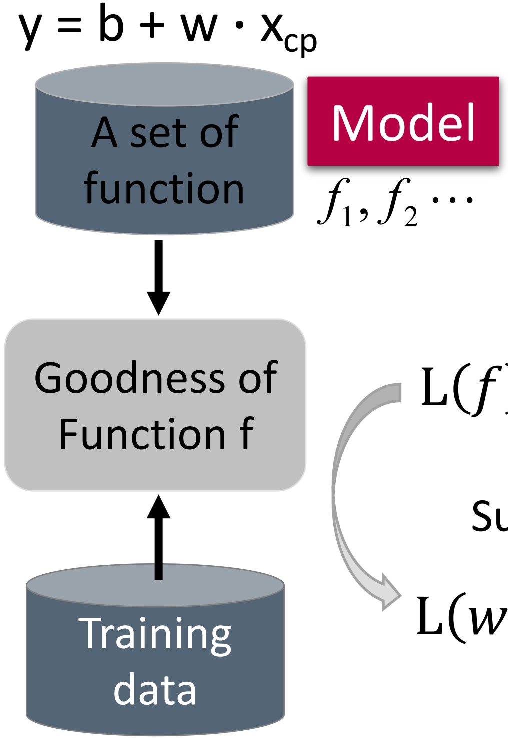
Training data

Function Input:

Function Output (scalar):



Step Two: Goodness of Function, Continued



Loss function L :

Input: A function

Output: How bad it is

Estimation error

$$\mathrm{L}(f) = \underbrace{\sum_{n=1}^{10}}_{\text{Sum over examples}} \boxed{\left(\hat{y}^n - \underbrace{f\left(x_{cp}^n\right)}_{\text{Estimated y based on input function}}\right)^2}$$

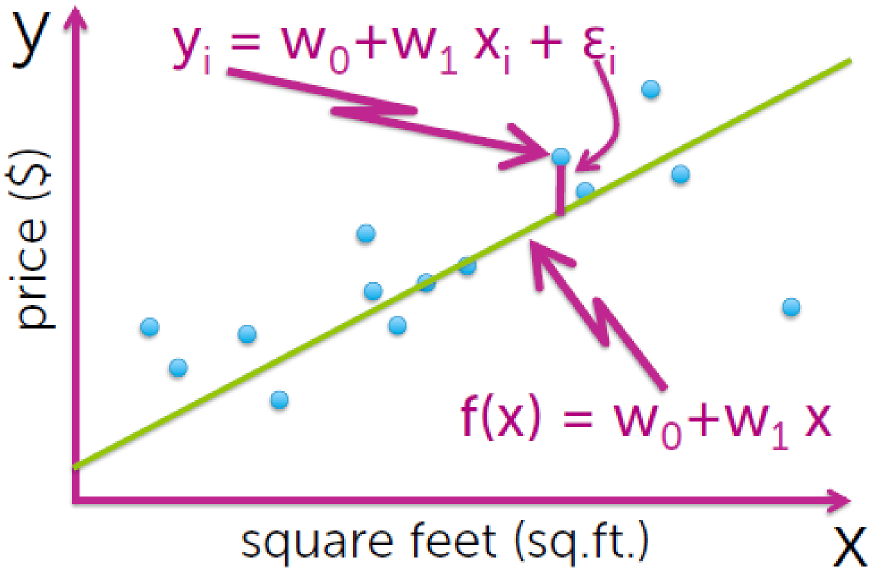
Sum over examples

Estimated y based on input function

$$\mathrm{L}(w, b) = \sum_{n=1}^{10} \left(\hat{y}^n - \left(b + w \cdot x_{cp}^n\right)\right)^2$$

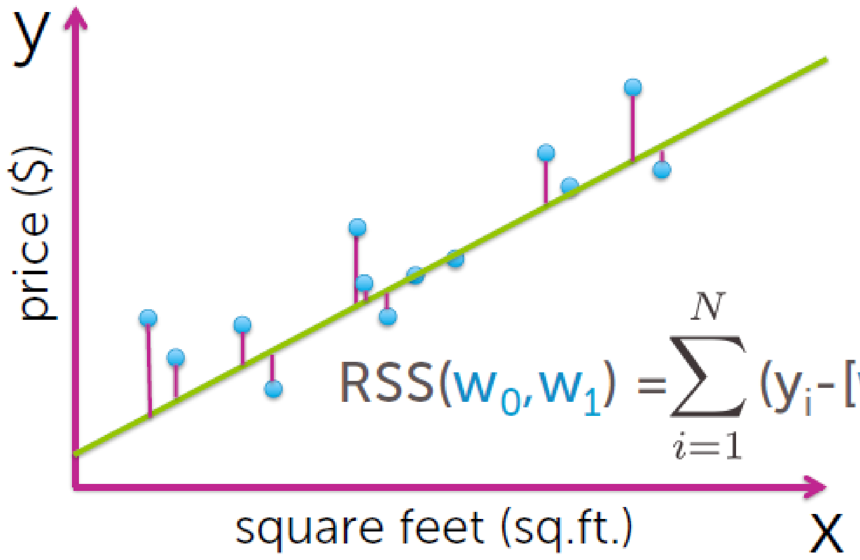
Simple Linear Regression Model

1 input and just fit a line to data



“Cost” of Using a Given Line

Residual sum of squares (RSS)



$$RSS(w_0,w_1) = \sum_{i=1}^{N} (y_i - [w_0 + w_1 x_i])^2$$

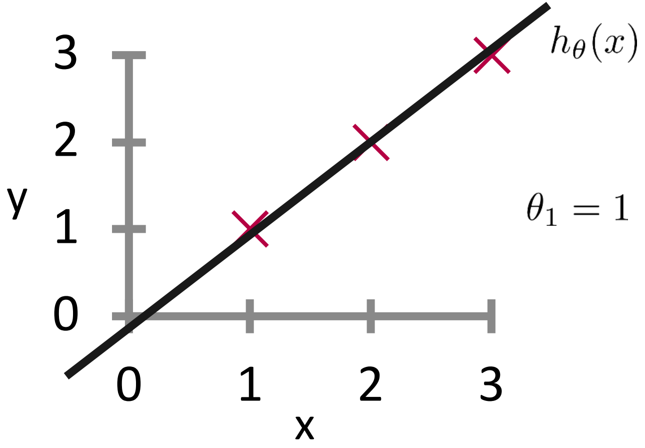
Example I: Part I

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

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

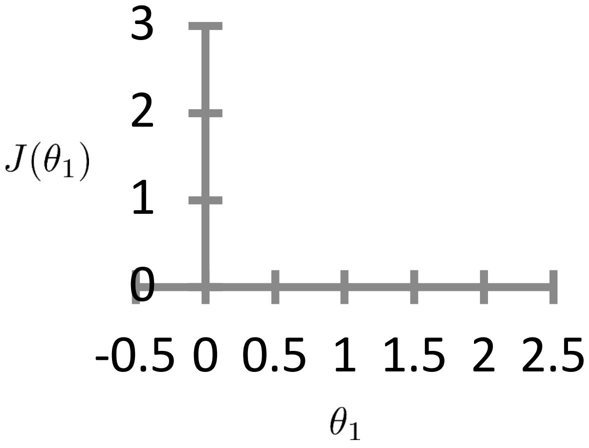
$h_\theta(x)$

(for fixed θ_1, this is a function of x)



$J(\theta_1)$

(function of the parameter θ_1)

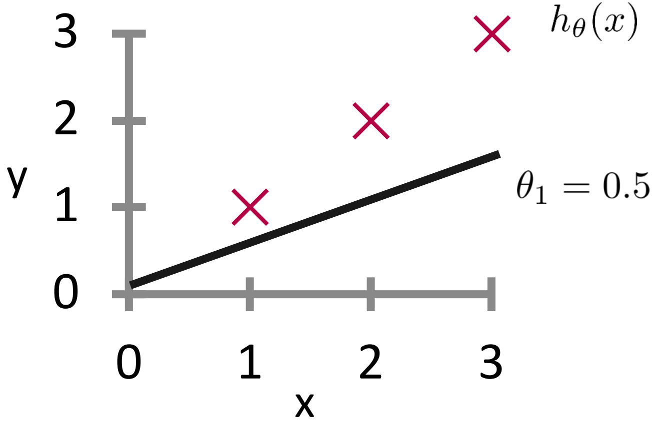


Example I: Part II

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

$$h_\theta(x)$$

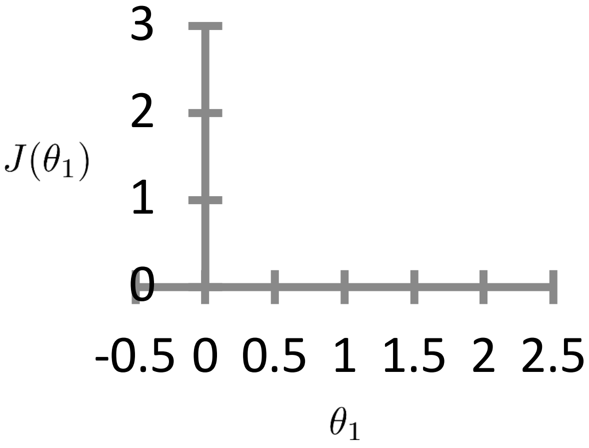
(for fixed θ_1, this is a function of x)



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

$$J(\theta_1)$$

(function of the parameter θ_1)



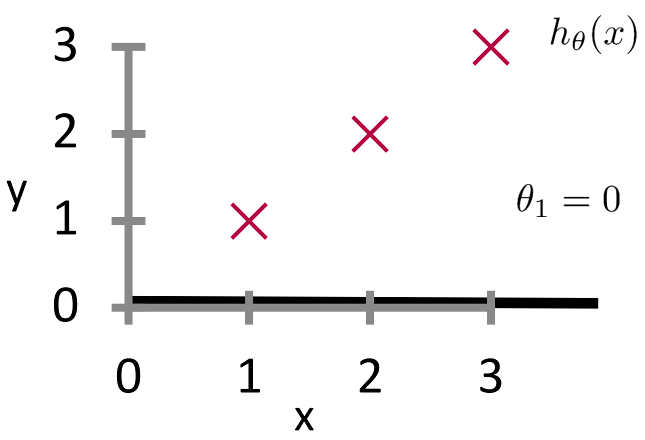
Example I: Part III

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

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

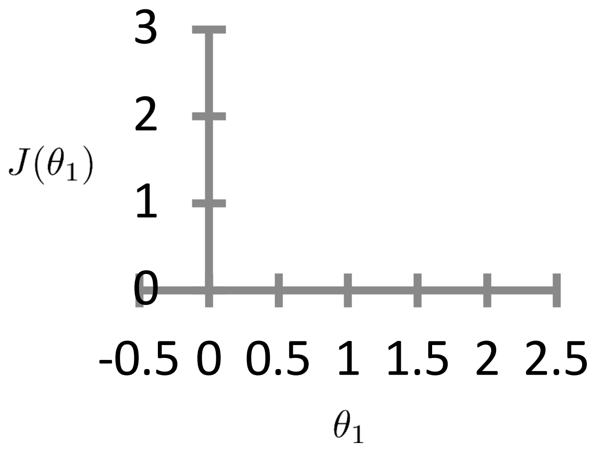
$h_\theta(x)$

(for fixed θ_1, this is a function of x)



$J(\theta_1)$

(function of the parameter θ_1)

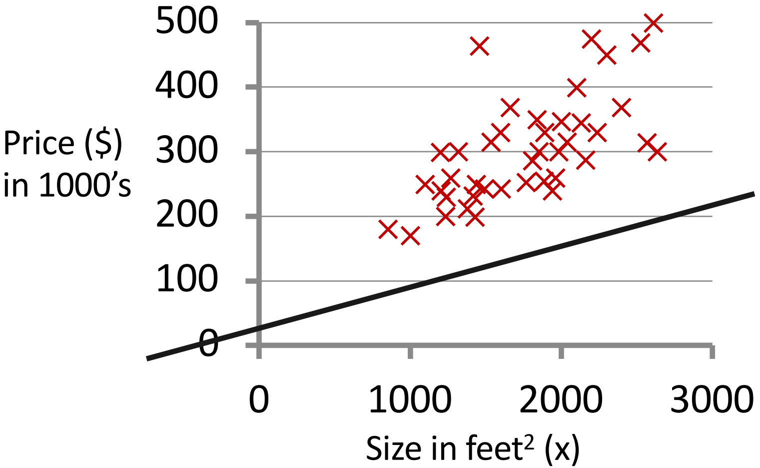


Example II: Part I

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

$$h_\theta(x)$$

(for fixed θ_0, θ_1, this is a function of x)



$$h_\theta(x) = 50 + 0.06x$$

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

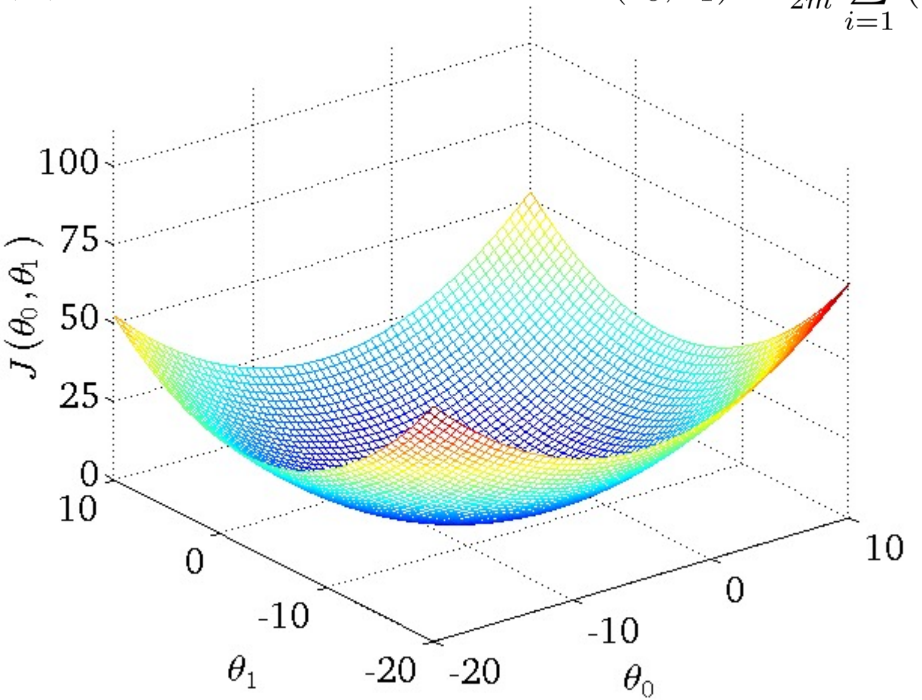
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)

Example II: Part II

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

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



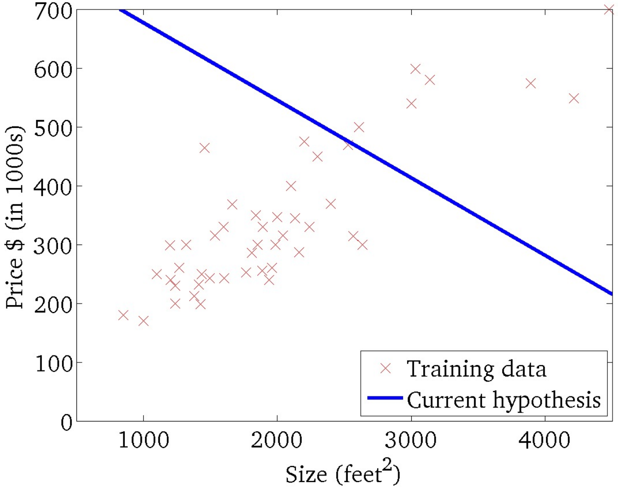
Example II: Part III

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

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

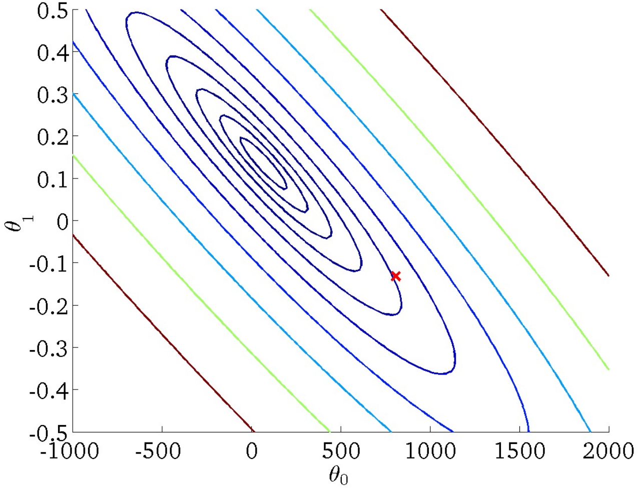
$$h_\theta(x)$$

(for fixed θ_0, θ_1, this is a function of x)



$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



Example II: Part IV

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

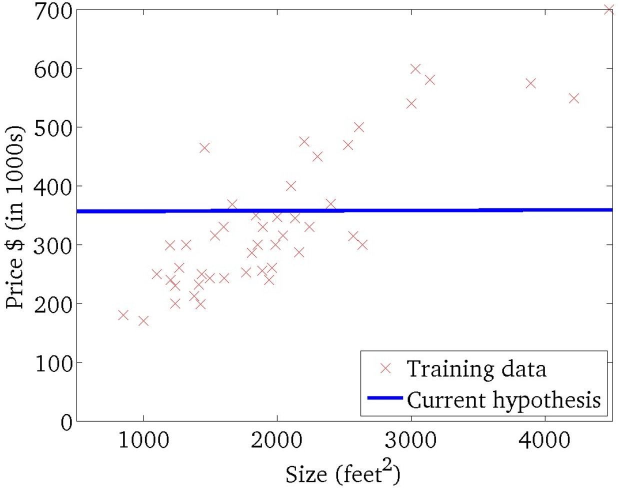
$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

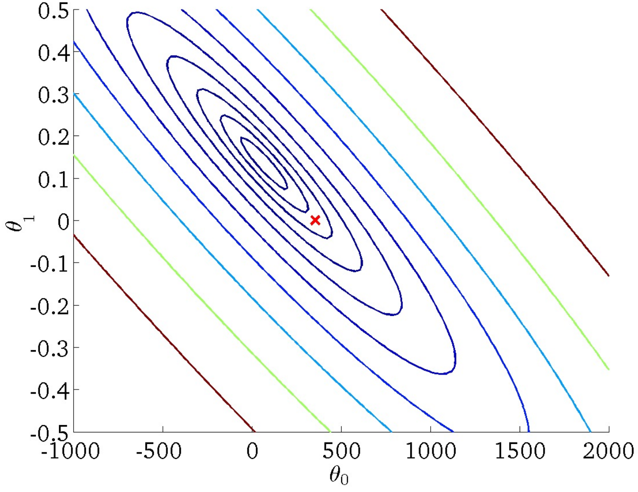
$$h_\theta(x)$$

$$J(\theta_0, \theta_1)$$

(for fixed θ_0, θ_1, this is a function of x)

(function of the parameters θ_0, θ_1)





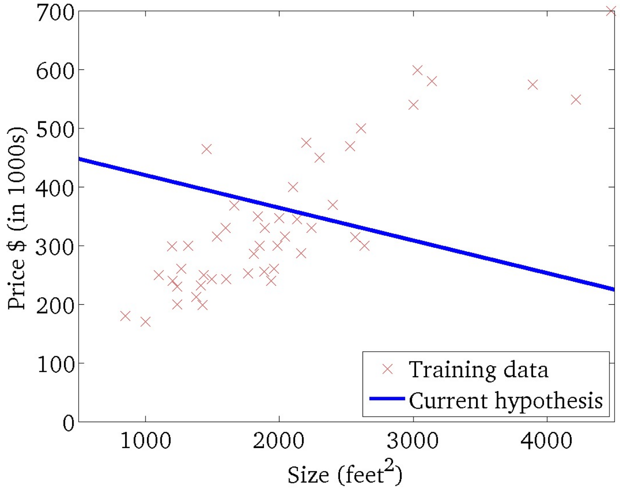
Example II: Part V

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

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

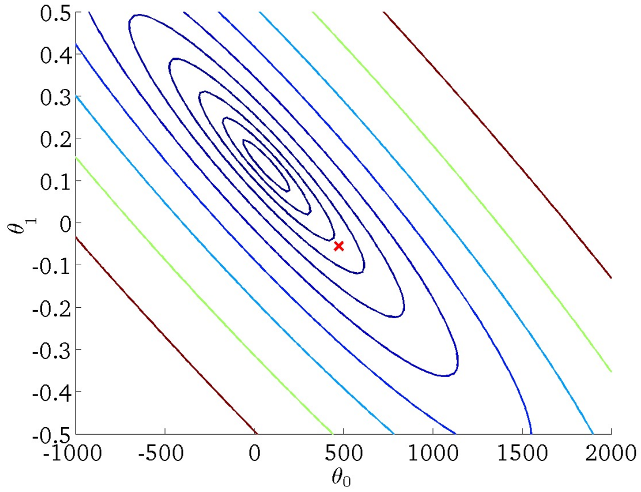
$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



Example II: Part VI

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

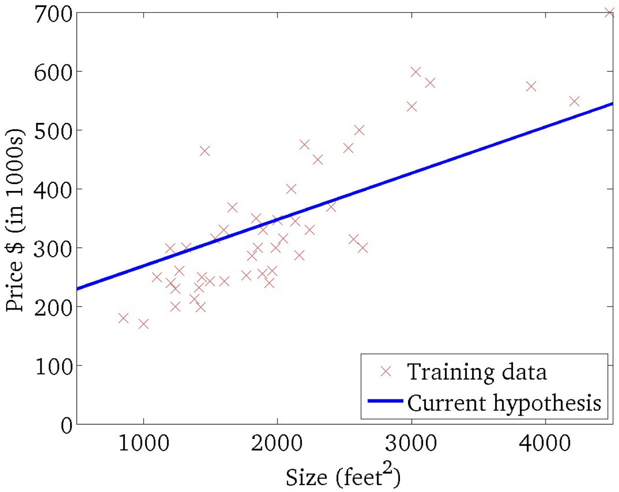
$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

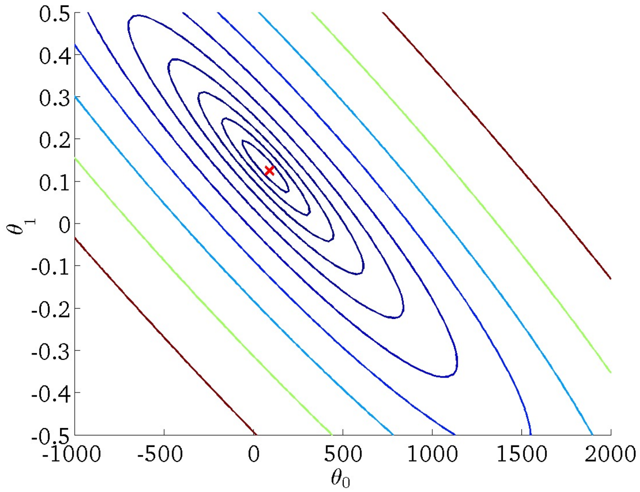
$$h_\theta(x)$$

$$J(\theta_0, \theta_1)$$

(for fixed θ_0, θ_1, this is a function of x)

(function of the parameters θ_0, θ_1)





Step Three: Best Function

A set of function

Model

$f_1, f_2 \cdots$

$$\mathrm{L}(w, b) = \sum_{n=1}^{10} \left(\hat{y}^n - (b + w \cdot x_{cp}^n)\right)^2$$

Goodness of function f

Training data

Pick the "Best" Function

$$f^* = arg \min_f L(f)$$

$$w^*, b^* = arg \min_{w,b} L(w, b)$$

$$= arg \min_{w,b} \sum_{n=1}^{10} \left(\hat{y}^n - (b + w \cdot x_{cp}^n)\right)^2$$

Gradient Descent