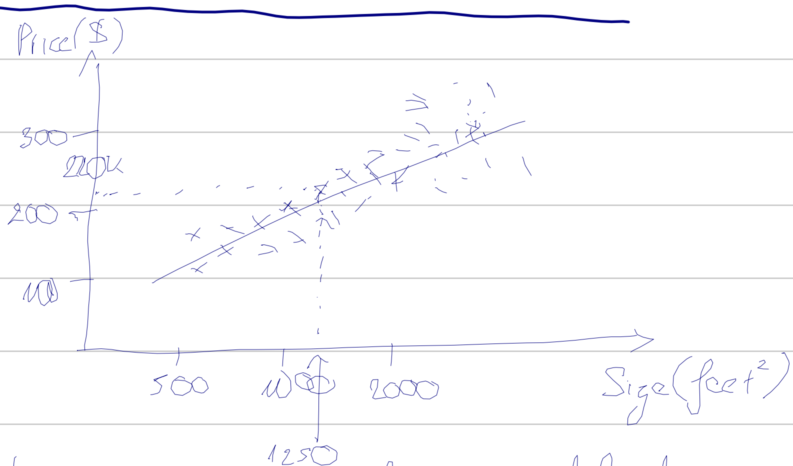
Linear Regression (with one variable).

Housing Prices
(Portland, OR)



Prediction of housing price based on model data set:

Supervised learning
Given the "right answer" for each example in the data

Regression Problem.
Predict real-valued output.

Classification: Discrete-valued output.

Training set

Size in feet^2 (x)	Price ($) in 1000's ($y$)
2104	460
1416	232
1534	315
...	...

$m = 47$

Notation: m = number of training examples

x's = "input" variable / features

y's = "output" variable ("target")

$(x, y) \rightarrow$ one training example

$(x^{(i)}, y^{(i)}) \leftarrow$ i^{th} training example

Training Set

↓

Learning Algorithm.

↓

Size of house → h → Estimated Price.

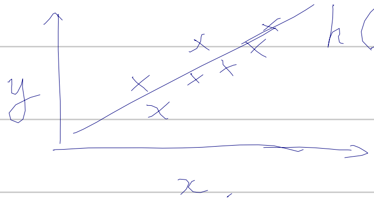
x hypotheses (estimated value of y)

h maps from x's to y's.

How do we represent h?

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

Shorthand: $h(x)$



$h(x) = \theta_0 + \theta_1 x$

Linear Regression with one var

Univariate linear regression.

↑ one variable

Cost function:

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$.

Idea: choose θ_0, θ_1 so that $h_\theta(x)$ is close to y for our training examples (x, y)

$\Rightarrow$ Minimize the output of the hypothesis & the actual value:

Minimize (θ_0, θ_1) $\quad (h_\theta(x) - y)^2$

size of training set $\rightarrow m$

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

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

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

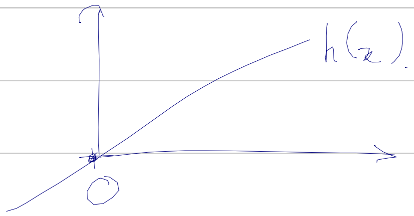
Goal: $\underset{\theta_0, \theta_1}{\text{Minimize}} \; \underbrace{J(\theta_0, \theta_1)}$

Cost function / Square error function

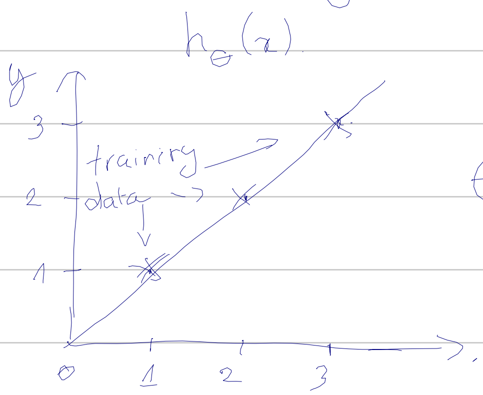
Simplified:

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

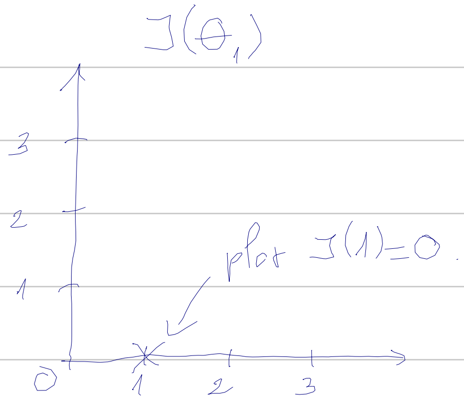
$\theta_0 = 0$ $\Rightarrow$



Assume training set $m = 3$, $\theta_1 = 1$



$\theta_1 = 1$



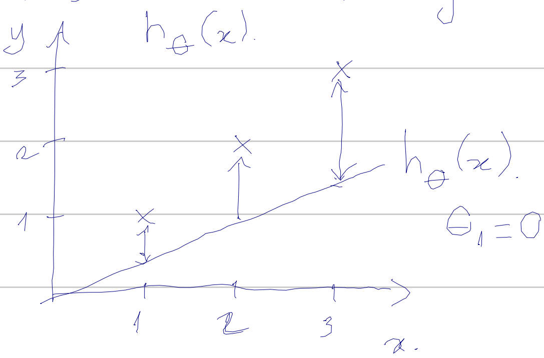
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_1 x^{(i)} - y^{(i)}\right)^2 = \frac{1}{2m}\left(0^2 + 0^2 + 0^2\right) = 0$$

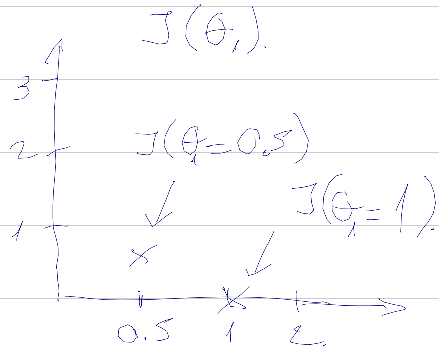
3 training data

$\Rightarrow J(1) = 0$.

Assume same training set, $\theta_1 = 0.5$

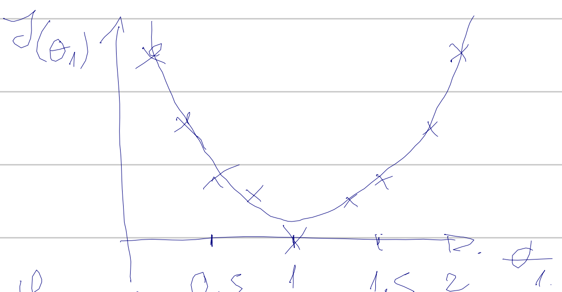


$\theta_1 = 0.5$



$$J(0.5) = \frac{1}{6}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right] \approx 0.58$$

$$J(0) = \frac{14}{6}$$



if we change θ_1 and continue to plot $J(\theta_1)$ $\Longrightarrow$ we get this diagram $\Longrightarrow$ $\boxed{\text{Our objective is to MINIMIZE } J(\theta_1)}$

Gradient Descend Algorithm

repeat until convergence {

$$\theta_j := \theta_j - \underbrace{\alpha}_{\text{learning rate}} \underbrace{\frac{\partial}{\partial \theta_j}}_{\text{derivative of } J} J(\theta_0, \theta_1)$$