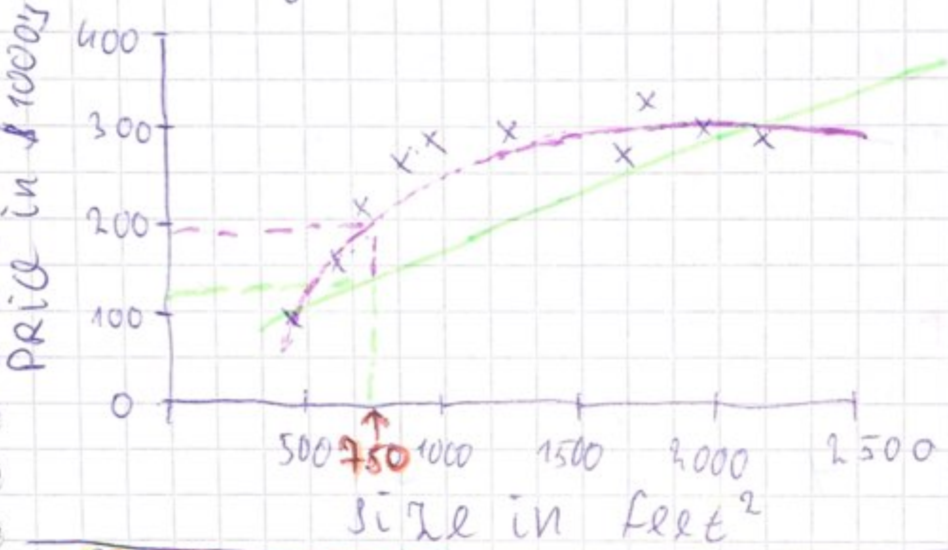
Machine learning
Week 1.
Supervised learning

Housing price predictions

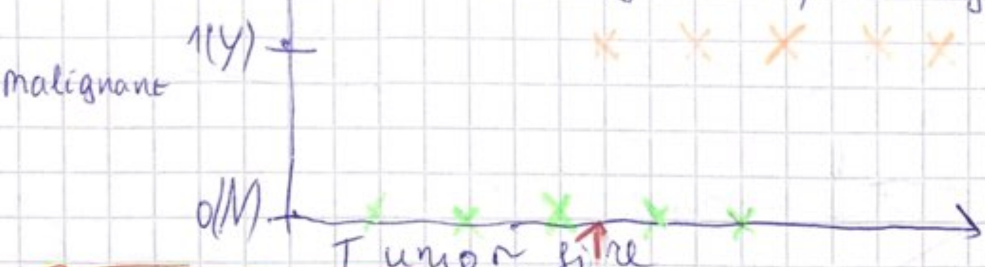


Supervised Learning

"right answers" given

Regression Predict continuous valued output (price)

Breast cancer (malignant, benign)

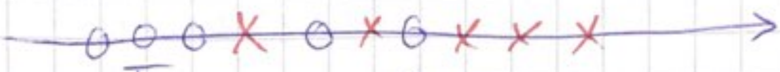


Classification Discrete valued output (0 or 1)

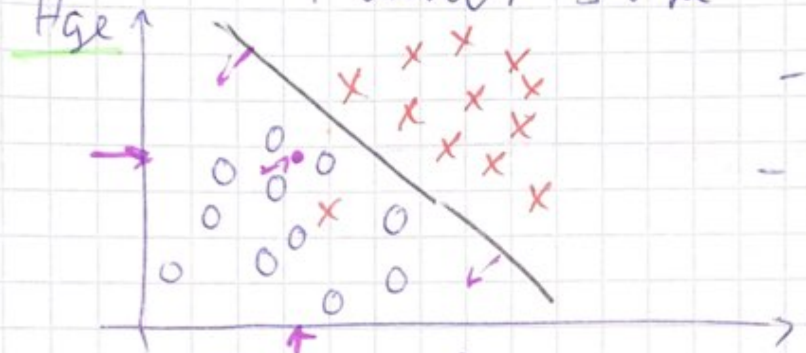
0, 1, 2, 3

↓ ↓ ↓ ↓

Benign (No Cancer), type 1, type 2, type 3



Tumor Size



- Clump Thickness
- Uniformity of cell shape/size

test

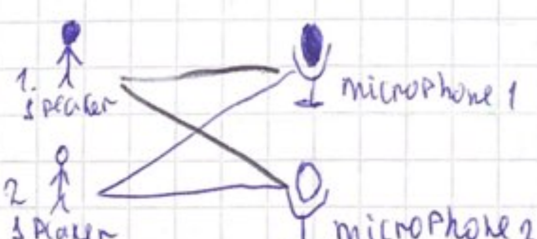
Pr.1 You have identical items.
Predict how many will sell over 3 month

Pr.2 Decide is account hacked/compromised

- Pr.1 - regression. Pr.2 - classification

Unsupervised Learning

- Clustering
- Coctail Party problem




```
[W,s,v] = svd((repmat(sum(x.*x,1),size(x,1),1).*x)*x');
```

Test

Which one you address using an Un. Lear.

− Email labelad as spam/not spam, learn a spam filter.

+ Group News into set of articles aboun same storyes. Set of articles founded on the Web

+ Discover market segments and group customers into different market segments

− Patients labelad diabetes/not, leart to classify diabetes/not

Unsupervised learning allow to approach problems with little or no idea what our result should look like.

With U.L. there is no feedback based on the prediction results

Model and Cost Function

Model representation

- Super. Learning
- Regression Pr.

Training set of Housing prices

Feet^2 (x)	$1000 (y)
2104	460
1416	232
1534	315
852	178
...	...

$m = 47$

Notation:

m = Number of training examples
x's = "input" variable / features
y's = "output" variable / "target" variable

(x, y) - one training example

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

$x^{(1)} = 2104$
$x^{(2)} = 1416$
$y^{(1)} = 460$

Test

What is $y^{(3)}$?

- 315

Training Set
↓
Learning Algorithm
↓
Size → h → Price

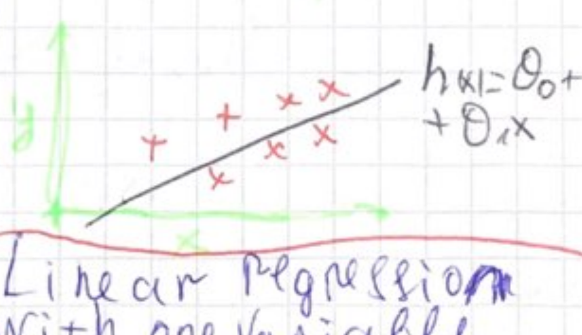
x — hypothesis — estimated value of y

h maps from x's to y's

How do we represent h?

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

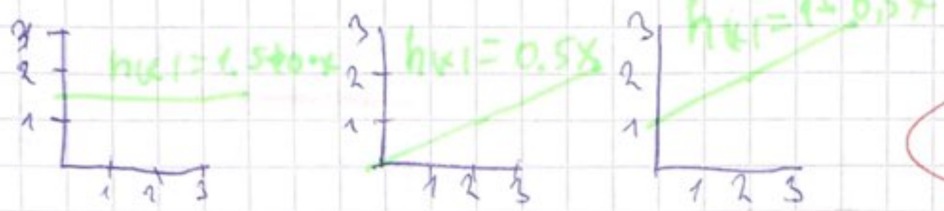
Shorthand: $h(x)$



Linear regression with one variable
Univariate linear regression.

Cost Function

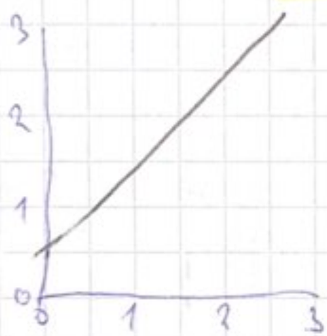
Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$



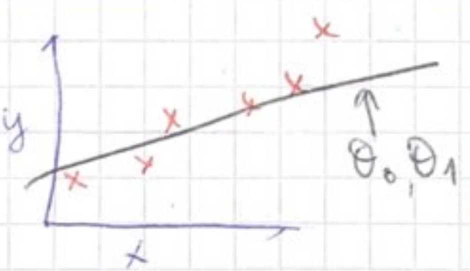
$h_\theta(x) = h(x)$

$\theta_0 = 1.5$	$\theta_0 = 0$	$\theta_0 = 1$
$\theta_1 = 0$	$\theta_1 = 0,5$	$\theta_1 = 0,5$

Test



- $-$ $\theta_0 = 0,\ \theta_1 = 1$
- $+$ $\theta_0 = 0,5,\ \theta_1 = 1$
- $-$ $\theta_0 = 1,\ \theta_1 = 0,5$
- $-$ $\theta_0 = 1,\ \theta_1 = 1$



Idea: Choose θ_0, θ_1 so that $h_\theta(x)$ is close to y for our tr. exam.

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

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

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

$\underset{\theta_0,\theta_1}{\text{minimize}}\ \underbrace{J(\theta_0,\theta_1)}_{\text{Cost function}}$

(squared error F.)

Cost function Intuition I

Hypothesis:

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

Parameters:

θ_0, θ_1

Cost function:

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

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

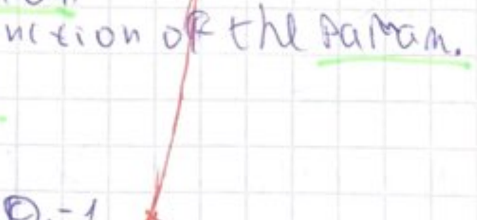
Simplified

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

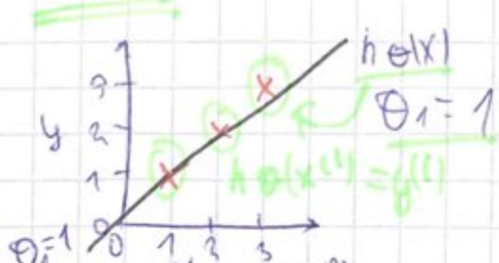
$\theta_0 = 0$

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

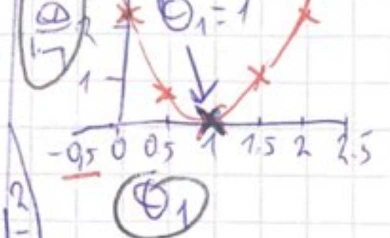
$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$



$h_\theta(x)$
(for fixed θ_1, this is a func. of x)



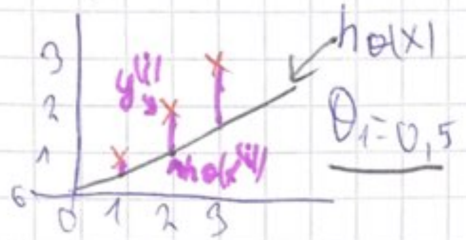
$J(\theta_1)$
(function of the param. θ_1)



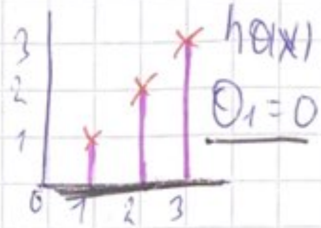
$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 =$

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

$\Theta_1 = 0,5$?



$\Theta_1 = 0$

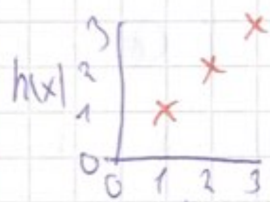


$J(0) = \frac{1}{2m}(1^2 + 2^2 + 3^2) = \frac{1}{6} * 14 \approx 2.3$

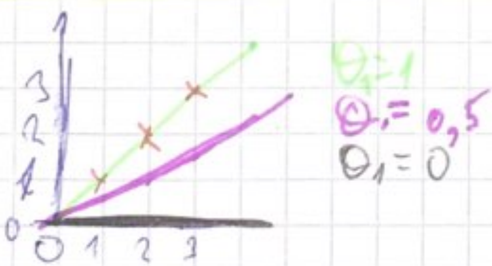
$$J(0,5) = \frac{1}{2m}\left[(0,5-1)^2 + (1-2)^2 + (1,5-3)^2\right] = \frac{1}{2\times3}(3,5) = \frac{3,5}{6} \approx \underline{0,58}$$

Test

We have a tr. set with $m=3$ examples $h_\theta(x) = \theta_1 x$. What is $J(0)$



- 0
- 1/6
- 1
+ 14/6



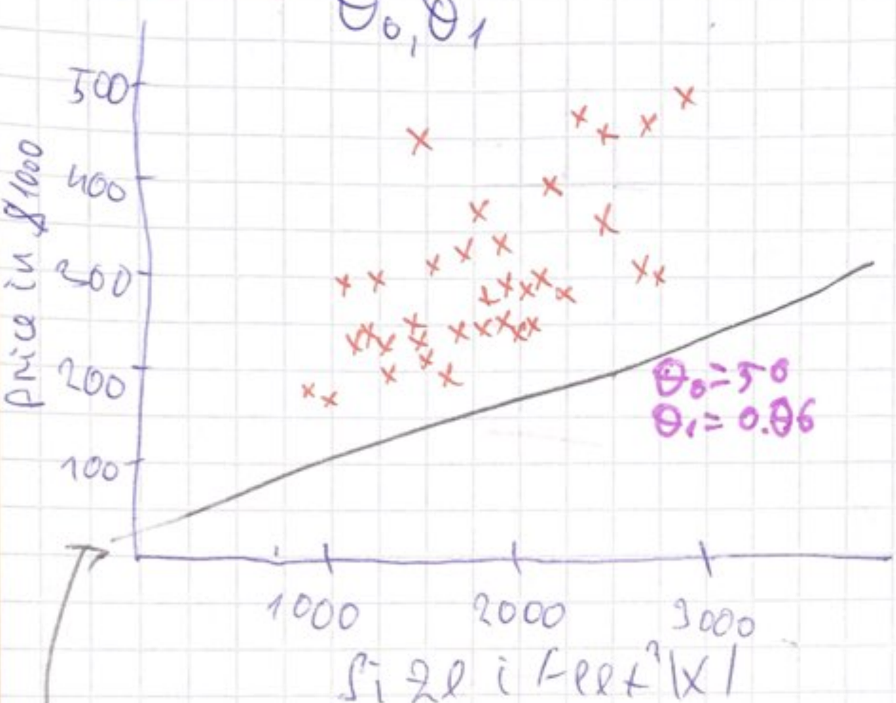
Cost function intuition II

Hypothesis: $h_\theta(x) = \Theta_0 + \Theta_1 x$

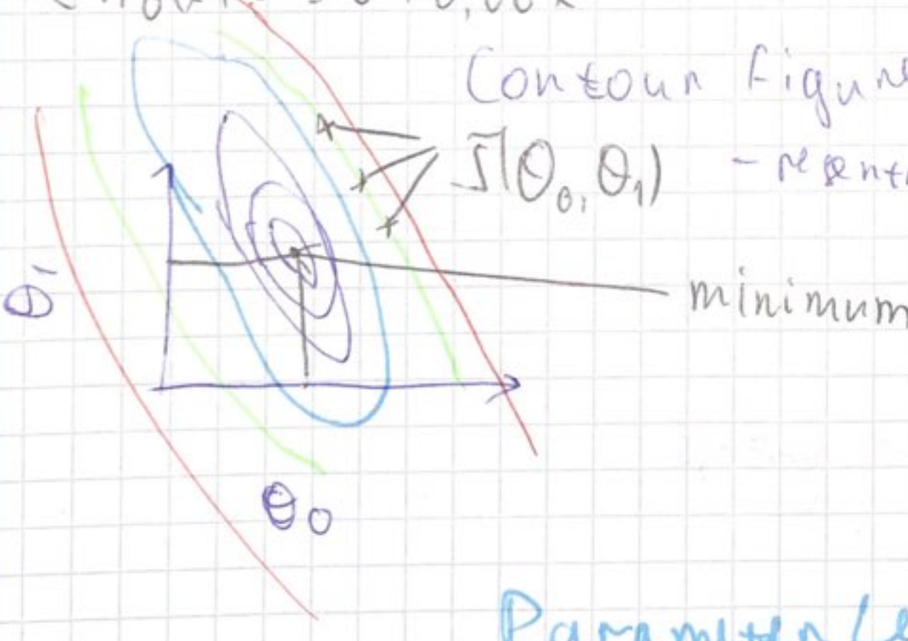
Paramete: Θ_0, Θ_1

Cost Function: $J(\Theta_0, \Theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$

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



$h_\theta(x) = 50 + 0{,}06x$



Contour figures - 3d representation in 2d $J(\theta_0, \theta_1)$

Parameter Learning

Gradient descent

Have some function $J(\theta_0, \theta_1)$ $J(\theta_0, \theta_1, \ldots, \theta_n)$

Want $\underset{\theta_0, \theta_1}{\min} J(\theta_0, \theta_1)$ $\underset{\theta_0, \ldots \theta_n}{\min}$

Outline

- Start with some θ_0, θ_1 (say $\theta_0 = 0$, $\theta_1 = 0$)
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ until we hopefully end up at a minimum

assignment $a := b$ ~~truth assertion $a = b$~~

Gradient descent algorithm

repeat until convergence {

$\boxed{\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$ (for $j = 0$ and $j = 1$)

}

learning rate — derivative — simultaneous update θ_0 and θ_1

Correct: Simultaneous update

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

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

Test

$\theta_0 = 1,\ \theta_1 = 2\ ;\ \theta_j := \theta_j + \sqrt{\theta_0 \theta_1}$

resulting value of θ_0 and θ_1?

- $\theta_0 = 1, \theta_1 = 2$
+ $\theta_0 = 1 + \sqrt{2}, \theta_1 = 2 + \sqrt{2}$
- $\theta_0 = 2 + \sqrt{2}, \theta_1 = 1 + \sqrt{2}$
- $\theta_0 = 1 + \sqrt{2}, \theta_1 = 2 + \sqrt{(1+\sqrt{2}) \cdot 2}$

Gradient descent intuition

rep. until convergence {

$\theta_j := \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$

learning rate

derivative

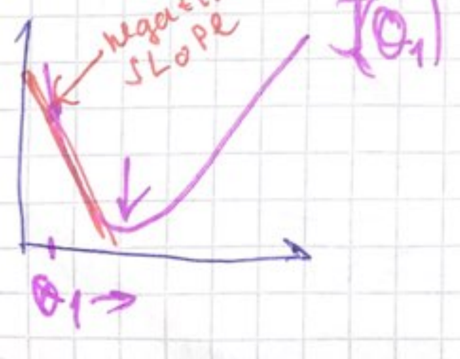
}

$\min_{\theta_1} J\theta_1 \quad \theta_1 \in \mathbb{R}$ (real number)

$\frac{\partial}{\partial \theta_1}$

$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J\theta_1$

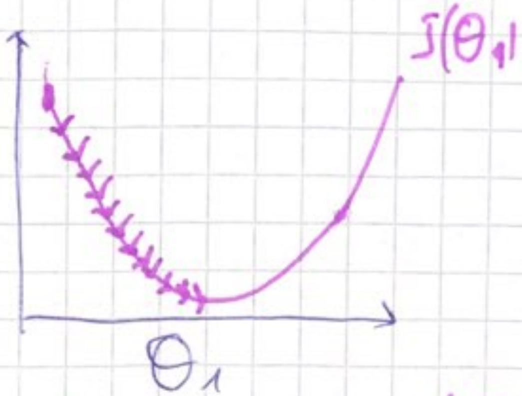
$\theta_1 := \theta_1 - \alpha \cdot$ (positive number)



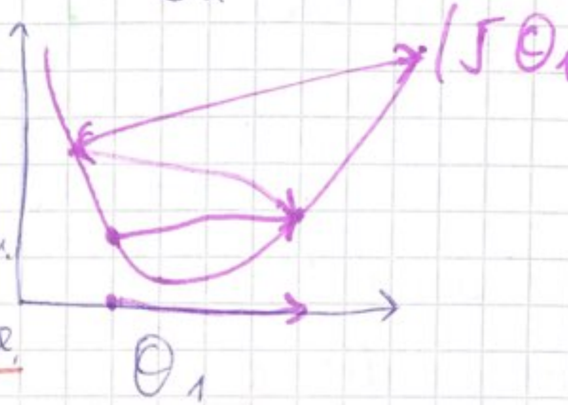
$\frac{d}{d\theta_1} J(\theta_1)$

≤ 0

$\theta_1 := \theta_1 - \alpha$ (negative number) if α is too small, gradient descent can be slow

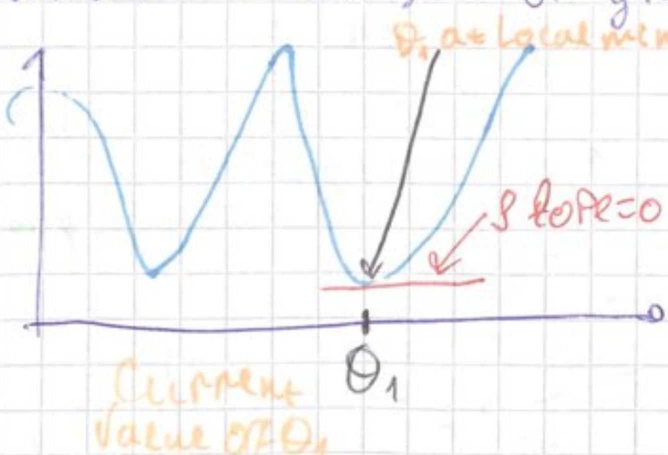


if α is too large, gr. descent can overshoot the minimum. It may fail to converge or even diverge.



test

Suppose θ_1 is at local optimum of $J(\theta_1)$. What will one step of g.d. do?



+ θ_1 unchanged
- Change θ_1 in random direction
- Move θ_1 to global minimum
- Decrease θ_1

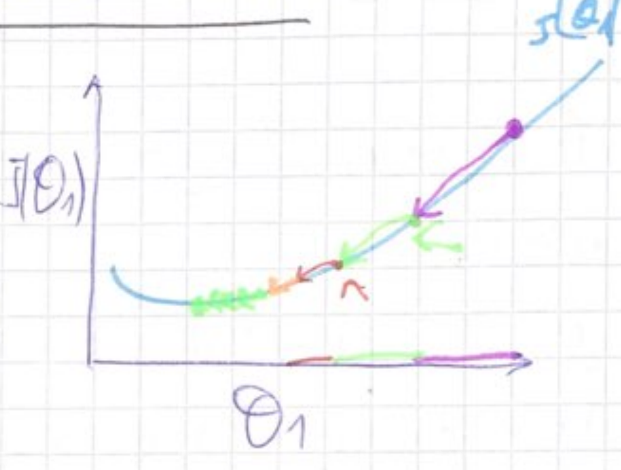
$$\theta_1 := \theta_1 - \alpha \underbrace{\boxed{\frac{d}{d\theta_1} J(\theta_1)}}_{=0}$$

$\theta_1 := \theta_1 - \alpha \cdot 0$

$\theta_1 := \theta_1$

Gradient descent can converge to a local minimum, even the learning rate α fixed.

$$\theta_1 := \theta_1 - \alpha \boxed{\frac{d}{d\theta_1} J(\theta_1)}$$



As we approach a local minimum grade. Will atomatically take smaller steps. So no need to decrease α over time

Gradient descent for linear regression

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

$$= \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)^2$$

$$\theta_0 \; j = 0 : \frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

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

Derivatives

Gradient descent algorithm

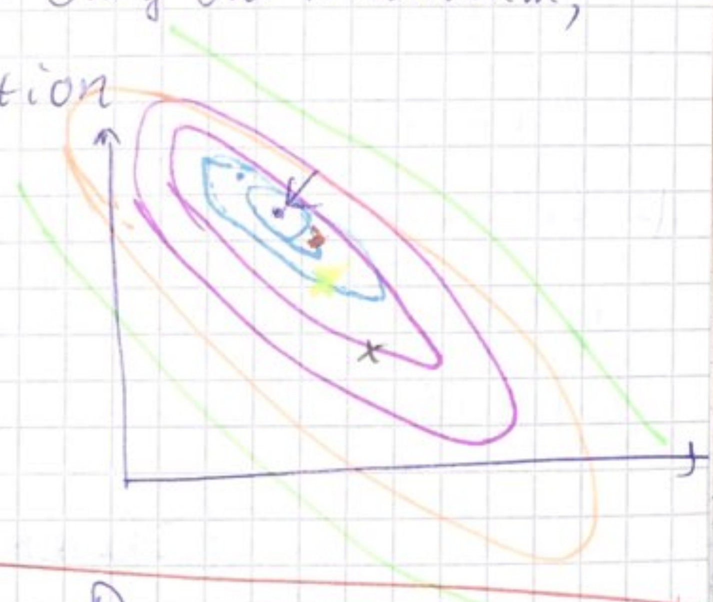
repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \boxed{\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})} \quad \frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \boxed{\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x^{(i)}} \quad \frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1)$$

}

Convex function - only one minimum, Bow shaped function



"Batch" Gradient Descent

"Batch": Each step of grad. uses all the training examples.

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

test

What is true?

- To make gr. de. converge, we must slowly decrease α over time.

- Gr. de. is guaranteed to find the global minimum for any $J(\theta_0, \theta_1)$

+ Gr. de. can converge even α is kept fixed. (But α can't be too large, or else it may fail to converge)

+ For the specific choise of cost function $J(\theta_0, \theta_1)$ used in l. r, there are no local optima (other than global optimum).

Linear algebra review

Matrices and Vectors

$\boxed{\text{Matrix: Rectangular array of numbers}}$

$$\begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 944 & 1437 \\ 147 & 1448 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

4 × 2 matrix

2 × 3 matrix

$\boxed{\mathbb{R}^{4\times 2}}$

Dimension of matrix = number of row * number of column test

What is true:

$\begin{bmatrix} 1 & 2 \\ 4 & 0 \\ 0 & 1 \end{bmatrix}$ is 3×2 m. +

$\begin{bmatrix} 0 & 1 & 4 & 2 \\ 3 & 4 & 0 & 9 \end{bmatrix}$ is 4×2 m. −

$\begin{bmatrix} 0 & 4 & 2 \\ 3 & 4 & 9 \\ 5 & -1 & 0 \end{bmatrix}$ is 3×3 m. +

$\begin{bmatrix} 1 & 2 \end{bmatrix}$ is a 1×2 m. +

$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$

A_{ij} = "i, j entry" in the i^{th} row, j^{th} column

$A_{11} = 1402$ $\quad A_{41} = 147$

$A_{12} = 191$ $\quad A_{43} =$ undefined

$A_{32} = 1437$

Test

$A = \begin{bmatrix} 85 & 76 & 66 & 5 \\ 94 & 75 & 18 & 28 \\ 68 & 40 & 71 & 5 \end{bmatrix}$ What is A_{32}?

− 18 − 28 − 76 + 40

$\boxed{\text{Vector: An } n\times 1 \text{ matrix.}}$ $y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$ $\mathbb{R}^4$

4 dimension vector →

y_i = i^{th} element

$y_1 = 460 \quad y_3 = 315$

$y_2 = 232$

1-indexed vs 0-indexed

$$\boxed{y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}} \quad y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

y[1] 1-indexed

y[0] 0-index

Addition and Scalar Multiplication

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0,5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} = \boxed{\begin{bmatrix} 5 & 0,5 \\ 4 & 10 \\ 3 & 2 \end{bmatrix}}$$

3 × 2 matrix, 3 × 2 matrix, 3 × 2

$$\cancel{\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0,5 \\ 2 & 5 \end{bmatrix}} = \text{error}$$

3×2, 2×2

Test

$$\begin{bmatrix} 8 & 6 & 9 \\ 10 & 1 & 10 \end{bmatrix} + \begin{bmatrix} 3 & 10 & 2 \\ 6 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 11 & 16 & 11 \\ 16 & 2 & 9 \end{bmatrix}$$

Scalar Multiplication

real number

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 6 & 15 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} \times 3$$

3 × 2, 3 × 2

$$\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix} / 4 = \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3/2 & 3/4 \end{bmatrix}$$

Test

$$2 \times \begin{bmatrix} 4 & 5 \\ 1 & 7 \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 2 & 4 \end{bmatrix}$$

Combination of operation

m.addition/vectoraddition

$$3 \times \begin{bmatrix}1\\4\\2\end{bmatrix} + \begin{bmatrix}0\\0\\5\end{bmatrix} - \begin{bmatrix}3\\0\\2\end{bmatrix} / 3 = \begin{bmatrix}3\\12\\6\end{bmatrix} + \begin{bmatrix}0\\0\\5\end{bmatrix} - \begin{bmatrix}1\\0\\2/3\end{bmatrix} = \begin{bmatrix}2\\12\\10\frac{1}{3}\end{bmatrix}$$

Scalar mult.

Scalar division

3×1 m.
3dim.V.

Test

$$\begin{bmatrix}4\\6\\7\end{bmatrix} / 2 - 3\begin{bmatrix}2\\1\\0\end{bmatrix} = \begin{bmatrix}-4\\0\\3{,}5\end{bmatrix}$$

Matrix Vector multiplication

~~$1\times1 + 3\times5 = 16$~~

$$\begin{bmatrix}1 & 3\\4 & 0\\2 & 1\end{bmatrix}\begin{bmatrix}1\\5\end{bmatrix} = \begin{bmatrix}16\\4\\7\end{bmatrix}$$

3×2 2×1 3×1 matrix

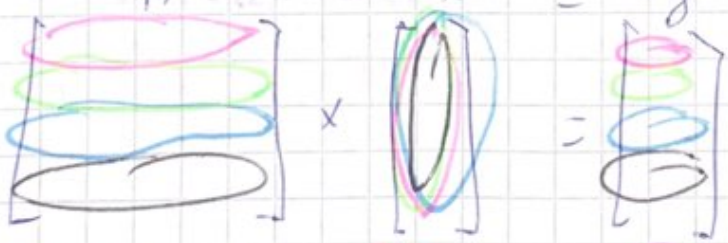
$1\times1 + 3\times5 = 16$

$4\times1 + 0\times5 = 4$

$2\times1 + 1\times5 = 7$

Details:

A × x = y



m×n matrix (m-rows n-columns)

n×1 matrix (n-dimensional vector)

m-dimensional vector

to get y_i multiply A's i^{th} row with elements of vector x, and add them up.

Test

$$\begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 3 & 0 & 4 \\ -1 & -2 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

What is dimension of the product

3×4 4×1 $\to 3\times 1 - 3\times 4 - 1\times 3 - 4\times 4$

$$\parallel$$

$$\begin{bmatrix} 14 \\ 13 \\ -7 \end{bmatrix}$$

$1\times 1 + 2\times 3 + 1\times 2 + 5\times 1 = 14$

$0\times 1 + 3\times 3 + 0\times 2 + 4\times 1 = 13$

$-1\times 1\ (-2)\times 3 + 0\times 2 + 0\times 1 = -7$

Test

$$\begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 3 & 1 & 2 \end{bmatrix}\times\begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 18 \\ 13 \end{bmatrix}$$

House sizes $h_\theta(x) = -40 + 0{,}25x$ $h_\theta(2104)$

2104
1416
1534
852

$$\underbrace{\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix}}_{4\times 2}\times\underbrace{\begin{bmatrix} -40 \\ 0{,}25 \end{bmatrix}}_{2\times 1} = \underbrace{\begin{bmatrix} -40\times 1 + 0{,}25\times 2104 \\ -40\times 1 + 0{,}25\times 1416 \\ -40\times 1 + 0{,}25\times 1534 \\ -40\times 1 + 0{,}25\times 852 \end{bmatrix}}_{4\times 1}$$

$\boxed{\text{Prediction} = \text{DataMatrix} \times \text{Parameters}}$

Matrix-matrix multiplication

$$\underbrace{\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix}}_{2\times 3}\underbrace{\begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix}}_{3\times 2} = \underbrace{\begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix}}_{2\times 2}$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix}\times\begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} 11 \\ 9 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix}\times\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ 14 \end{bmatrix}$$

$A \times B = C$

$[\ \ ] \times [\ \ ] = [\ \ ]$

m×n matrix, m-rows, n-columns; n×o matrix, n-rows, o-columns; m×o matrix

The i^{th} column of the matrix C is obtained by multiplying A with i^{th} column of B

for ($i = 1, 2, \ldots, o$)

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 7 \\ 15 & 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 1\times0+3\times3 \\ 2\times0+5\times3 \end{bmatrix} = \begin{bmatrix} 9 \\ 15 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1\times1+3\times2 \\ 2\times1+5\times2 \end{bmatrix} = \begin{bmatrix} 7 \\ 12 \end{bmatrix}$$

Test

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 7 & 9 \\ a & b \\ c & d \end{bmatrix}$$

What is a? -7 -12 +10 -6

What is b? -7 -10 +12 -15

What is c? -7 -12 +10 -15

What is d? -8 -10 -12 +15

House sizes have 3 competing hypotheses:

2104
1416
1534
852

1. $h_\theta(x) = -40 + 0.25x$
2. $h_\theta(x) = 200 + 0.1x$
3. $h_\theta(x) = -150 + 0.4x$

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix} = \begin{bmatrix} 486 & 410 & 692 \\ 314 & 342 & 416 \\ 344 & 353 & 464 \\ 173 & 285 & 191 \end{bmatrix}$$

prediction of 1 — prediction of 2 — prediction of 3

Matrix multiplication

properties.

$3 * 5 = 5 \times 3$ "Comutative"

Let A and B be matrix. Then in general
$A \times B \neq B \times A$ (not commutative

E.g. $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \neq$

$\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix}$

$\underset{m \times n}{A} \times \underset{n \times m}{B}$ is $m \times m$

$\underset{m \times n}{B} \times \underset{n \times m}{A}$ is $n \times n$

$3 \times 5 \times 2$ "Associative"

$3 \times 10 = 30$
15×2

$3 \times (5 \times 2) = (3 \times 5) \times 2$

$A \times B \times C$ $\quad A \times (B \times C) \leftarrow$
$(A \times B) \times C \leftarrow$

Let $D = B \times C$. Compute $A \times D$ $\quad A \times (B \times C)$

Let $E = A \times B$. Compute $E \times C$ $\quad (A \times B) \times C$

Same answer

Identity matrix $\quad$ 1 is identity

$1 * z = z \times 1 = z$
for any z

Denoted I (or $I_{n \times n}$)

$\underset{2 \times 2}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$ $\underset{3 \times 3}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}$ $\underset{4 \times 4}{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}$

informally $\begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix}$

For any matrix A

$$\underset{m\times n}{A} \times \underset{n\times n}{\boxed{I}} = \underset{m\times m}{\boxed{I}} \times \underset{m\times n}{A} = \underset{m\times n}{A}$$

Note: $AB \neq BA$ in general

$AI = IA$

Test:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

inverse and traspose

1 = "identity" $\quad 3(\underbrace{3^{-1}}_{1/3}) = 1 \quad 12 \times (\underbrace{12^{-1}}_{1/12}) = 1$

Not all number have an inverse.

$0(0^{-1})$ - undefined.

square m. r. = c.

Matrix inverse: if A is an $m \times m$ matrix, and if it has an inverse,

$AA^{-1} = A^{-1}A = I$ $\quad A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ no inverse!

Only square matrix have inverse

E.g. $\underset{\substack{2\times 2 \\ A}}{\begin{bmatrix} 3 & 4 \\ 2 & 16 \end{bmatrix}} \underset{A^{-1}}{\begin{bmatrix} 0.4 & -0.1 \\ -0.05 & 0.075 \end{bmatrix}} = \underset{A^{-1}A}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} = I_{2\times 2}$

Matrices that don't have an inverse are singular or degenerate

Matrix Transpose

$$A = \underset{2\times 3}{\begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix}} \quad B = A^T = \underset{3\times 2}{\begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 0 & 9 \end{bmatrix}}$$

Let A be an $m \times n$ matrix, and let $B = A^T$. Then B is an $n \times m$ matrix, and

$$B_{ij} = A_{ji}$$

$B_{12} = A_{21}$

$B_{32} = A_{23}$

Test

What is? $\begin{bmatrix} 0 & 3 \\ 1 & 4 \end{bmatrix}^T = \begin{bmatrix} 0 & 1 \\ 3 & 4 \end{bmatrix}$