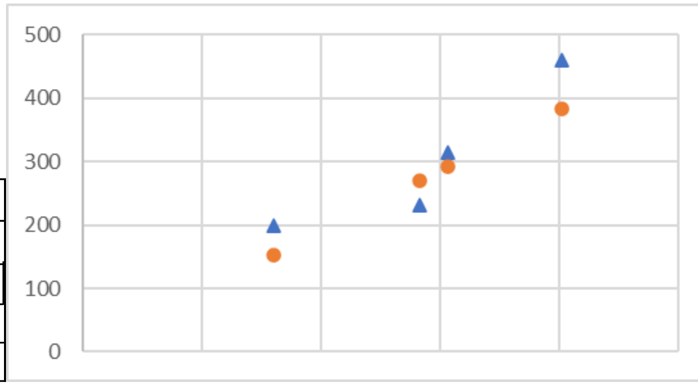
Exercise 1_ Prediction and inference_ big data_machine learning

1.- 1.- We want to predict the price of a new house in the USA. To run that prediction, we knowsome prices (Y dependent variable) and their corresponding surface (X feet ^2) (independent variable or predictor. We suppose a linear relationship between this $h_\Theta=\Theta_0+\Theta_1X$, being X the feet ^2 and h_Θ the prediction. To be able to have a prediction, we need to calculate Θ_0 and Θ_1 fitting the linear function to the data. In the graph below circles are the real prices and triangles the prediction

X	Y
feet ^2	price$
2013	460
1416	232
1534	315
800	200



The criterion to find what is the best possible solution is to minimise the summation of squared

errors using the formula (m is the size of the sample, in our case 4)

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

ITERATION 0

To find what is the minimum error we begin guessing some solution Θ0=0 and

Θ1=0,21, and we calculate the summation of the square error (please fill in the blanks)

Θ_0	0
Θ_1	0,19

feet ^2	price$	h_Θ	h_Θ-y_i	(h_Θ-y_i)^2
2013	460	382,47	-77,53	6010,90
1416	232			
1534	315	291,46	-23,54	554,13
800	200	152,00	-48,00	2304,00
		suma	-112,03	10240,99
			$J(\Theta_0,\Theta_1)$	

number of trair m 4

linear regression $h_\Theta=\Theta_0+\Theta_1X$

Then we calculate the new Θ0 and Θ1. To choose this new parameters we consider the previous solution, and we apply the algorithm called gradient descent. Lets be alfa the learning parameter

	j=0	j=1
feet ^2	$h_{\Theta 1}-y_i$	$(h_{\Theta 1}-y_i)*x_i$
2013	-77,53	-156067,89
1416		
1534	-23,54	-36110,36
800	-48	-38400
sum/m		
	alfa	0,00000001
	iter1	
	Θ_0	
	Θ_1	

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

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

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

- To calculate this fill in the blanks in the table and obtain the new Θ_0 and Θ_1

ITERATION 1

sing the las Θ0 and Θ1 calculated in the table above do the process again and calculate the iteration 2 parameters

2,80075E-07

X	Y			
feet ^2	price$	h_Θ	$h_\Theta-y_i$	$(h_\Theta-y_i)$^2
2013	460	383,37	-76,63	5872,70
1416	232	269,67	37,67	1419,07
1534	315	292,14	-22,86	522,44
800	200	152,36	-47,64	2269,93
		suma	-109,46	10084,14
			$J(\Theta_0,\Theta_1)$	1.261

number of trair m 4

linear regression $h_\Theta=\Theta_0+\Theta_1 X$

	j=0	j=1
	$h_{\Theta 1}-y_i$	$(h_{\Theta 1}-y_i)*x_i$
2013	-76,63	-154263,36
1416	37,67	**53341,54**
1534	-22,86	-35062,44
800	-47,64	-38114,99
sum/m		
	alfa	0,00000001
	iter 2	
	Θ_0	5,53734E-07
	Θ_1	

ITERATION 2

		X	Y			
5,53734E-07	feet ^2		price$	h_Θ	$h_\Theta - y_i$	$(h_\Theta - y_i)$^2
		2013	460	384,24	-75,76	5739,18
		1416	232			
		1534	315	292,81	-22,19	492,36
		800	200	152,70	-47,30	2236,87
				suma	-106,96	9934,30
number of trair m		4			$J(\Theta_0,\Theta_1)$	
linear regression	$h_\Theta=\Theta_0+\Theta_1 X$					

	j=0	j=1
feet ^2	$h_{\Theta 1} - y_i$	$(h_{\Theta 1} - y_i) * x_i$
2013	-75,75740768	-152499,6617
1416		
1534	-22,18920174	-34038,23548
800	-47,29554171	-37836,43337
sum/m	-26,7388151	-42540,0233
	alfa	0,00000001
	iter3	
	Θ_0	**8,21122E-07**
	Θ_1	