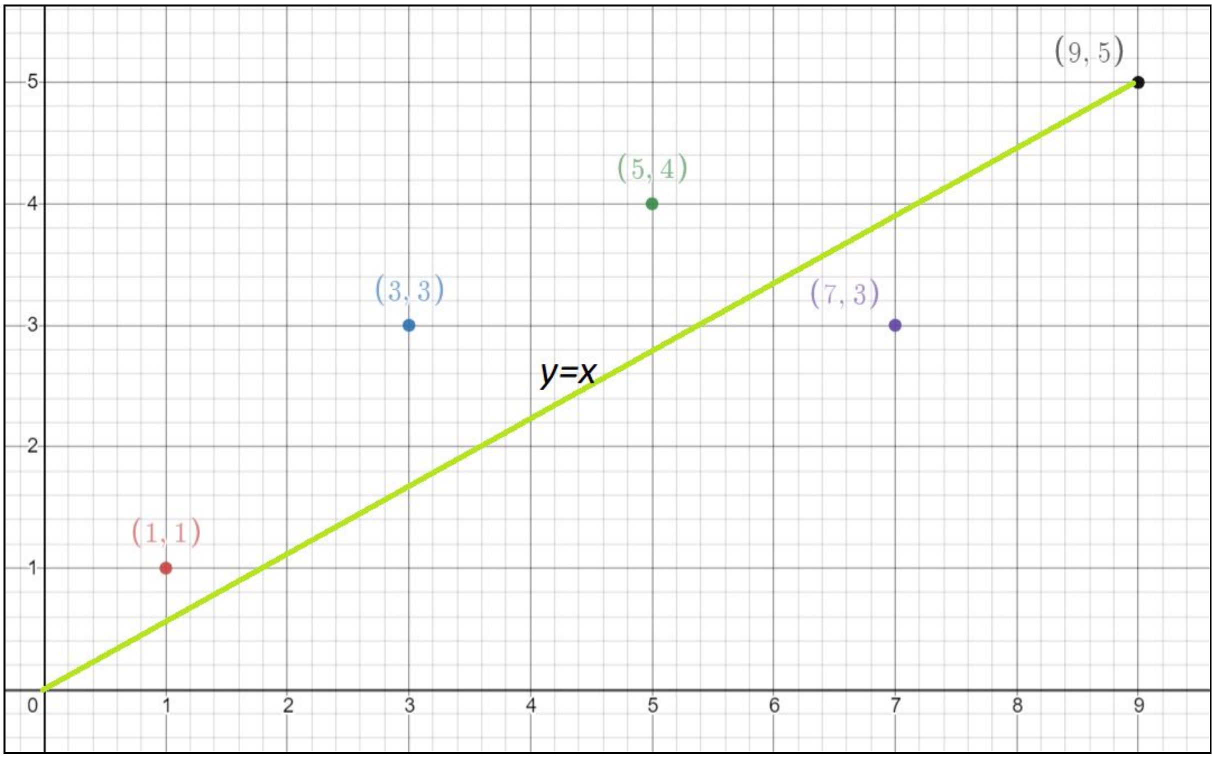
Assignment 1.2

Consider the error function on slide 29. Assume that you are trying to learn a linear function to approximate the true function. As training examples, the following points are observed: $(1,1)$, $(3,3)$, $(5,4)$, $(7,3)$, $(9,5)$.

a) Learn the function by using the LMS algorithm ($\eta = 0.1$). Only present each example once, in the order given by the above list. As initialization use the following linear function: $y = x$.



Since, we are given ***y = x*** that means, our intercept (w_0)=0 , so we have initial settings as $\mathbf{w_1=1}$, $\mathbf{w_0 = 0}$

#	x_i	y_i	$\hat{y}$ = $\mathbf{w_1 x_i}$	∇E	$\mathbf{w_0}$	$\mathbf{w_1}$
1	1	1	**0**+**1***1=1	1-1=0.0	0+(0.1*1*0.0)=0	1+(0.1*1*0.0)=1
2	3	3	0+1*3=3	3-3=0.0	0+(0.1*3*0.0)=0	1+(0.1*3*0.0)=1
3	5	4	0+1*5=5	4-5=-1.0	0+(0.1*5*(-1.0))=-0.5	1+(0.1*5*(-1.0))=0.5
4	7	3	-0.5+0.5*7=3.0	3-3=0.0	-0.5+(0.1*7*(0.0))=-0.5	0.5+(0.1*7*(0.0))=0.5
5	9	5	-0.5+0.5*9=4.0	5-4=1.0	-0.5+(0.1*9*(1.0))=0.4	0.5+(0.1*9*(1.0))=1.4

b) If all 5 training examples were given in advance, how can the best approximated linear function be directly calculated? What is it? (Hint: Refresh your mathematical knowledge on determining the minimum of a function, as you want to minimize the error function. You do not need to calculate everything by hand.)

We have the equation for LMS as, $\frac{\sum_{i=1}^{n}\left(y_i - y^{pred}\right)^2}{n}$

To find the value for coefficients we differentiate it wrt to β_0 & β_1

So we obtain the following equations for β_1 & β_0 respectively and for estimating the minimum of a function we equate to 0,

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\beta_0 \sum_{i=1}^{n} 1 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

So we estimate values from the data we have,

#	x_i	y_i	$x_i y_i$	x_i^2
1	1	1	1	1
2	3	3	9	9
3	5	4	20	25
4	7	3	21	49
5	9	5	45	81
Σ	25	16	96	165

Now we substitute these values in our equation and obtain,

β_1(165) + β_0(25) = 96 -------1 β_1(25) + β_0(5) = 16 ------2

So solve by basic linear algebra, we obtain $\beta_1 = \frac{2}{5}$ and by re-substituting either to any of above equations we estimate $\beta_0 = \frac{6}{5}$

c) Compare both results. What do you observe? Explain.

In a) we have to learn 2 times the weights for each of the samples. So this will increase if our training set (number of samples increases) and thereby increase our learning time.

In b) LMS algorithm provides the set of weights that minimizes the overall residuals, which can be later used for estimating the best fit line.