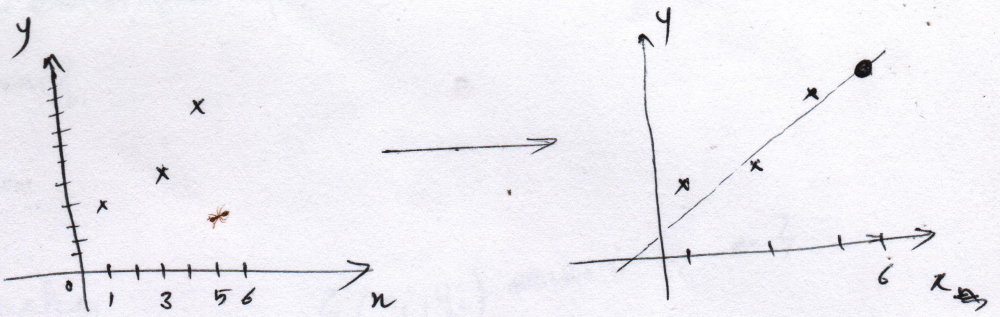
Linear Regression

x_n	y_n
1	4.8
3	11.3
5	17.2
6	?



We want to fit a line that closely passes through the given point. i.e We want to "linearly model the data"

So, the model has the form $y = f(x) = \theta_0 + \theta_1 x$.

Here, this is an example of Univariate linear regression. the input is scalar.

How do we find the parameters of the line (θ_0, θ_1) that closely agrees with the data?

(for least square approximation)

There is actually a analytical closed form solution to this problem.

(We will not be so fortunate in the upcoming topics, and we will have to resort to numerical optimization techniques).

Closed form Solution

$$\theta_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2} \quad \text{---(I)}, \qquad \theta_0 = \bar{y} - \theta_1 \cdot \bar{x} \quad \text{---(II)}$$

the bar (−) above the variable denotes the average value across the samples.

i.e $\bar{x} = \frac{x_1 + x_2 + x_3}{3}$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

x_n	y_n	$x_n \cdot y_n$	x^2
1	4.8	4.8	1
3	11.3	33.9	9
5	17.2	86	25
$\bar{x}=3$	$\bar{y}=11.1$	$\overline{xy}=41.57$	$\overline{x^2}=11.67$

Putting values into (I) & (II), we get

$$\theta_1 = \frac{41.57 - 3 \times 11.1}{11.67 - 3 \times 3} = \frac{8.27}{2.67} = 3.1$$

$$\theta_0 = 11.1 - 3.1 \times 3 = 1.8$$

Here. $y = f(x;\theta) = 1.8 + 3.1x$.

$$f(6) = 20.4$$

The solution obtained from the closed form Eqn. look reasonable. But where did this magic formula come from?.

Let's get to the details below

Least Square Approximation

Given a dataset $D = \{(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)\}$, We are going to attempt to find a functional relationship using a linear model defined as $f(x; \theta_0, \theta_1) = \theta_0 + \theta_1 x$. (remember, $y = mx + c$ 😊). this is the same equation. except we replaced the slope; m, by θ_1 and intercept, c, by θ_0.

For now, we have decided to define a least square cost function.

A cost function is a way to measure if our solution is good enough. (i.e the values of the parameter θ_0 & θ_1).

If the values of parameter θ_0 & θ_1 are chosen well, then it should fit the data well.

There can be many reasonable ways to specify this ξ.

for, example. 1

① Mean absolute difference (MAD) or L1-norm

$$L(f(x; \theta_0, \theta_1), D) = \sum_{i=1}^{N} |f(x_i; \theta_0, \theta_1) - y_i|.$$

② Mean Squared Error (MSE) or L2-norm.

$$L(f(x; \theta_0, \theta_1), D) = \sum_{i=1}^{N} (f(x_i; \theta_0, \theta_1) - y_i)^2.$$

Please Verify that both of these functions, have a small value when $f(x_i; \theta_0, \theta_1) \approx y_i$ and a large value when $f(x_i; \theta_0, \theta_1)$ is very different from y_i.