Simple Linear Regression Model:

$$y = \theta_0 + \theta_1 x + \varepsilon$$

where ε is a random error which has a Normal (Gaussian) distribution with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$

In practice, θ_0 and θ_1 are unknown, and need to be estimated using the sample data.

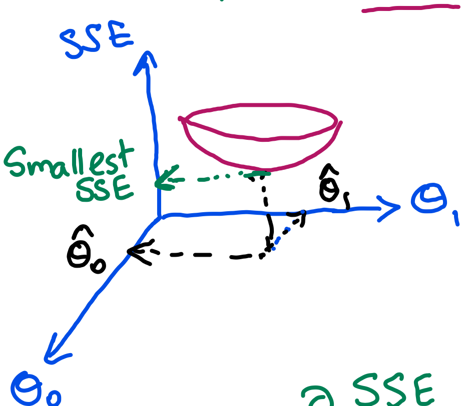
Let (x_i, y_i) be the ith pair of data and ε_i is the ith error for $i = 1, 2, \ldots, n$

Least Squares Method to estimate regression coefficients

ith error $= \varepsilon_i = y_i - [\theta_0 + \theta_1 x_i]$

(*) The sum of ε_i is Zero.

Consider $\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [y_i - (\theta_0 + \theta_1 x_i)]^2$ and this is called the Sum of Squared Error (SSE).



The Least Squares Method Find the values of θ_0 and θ_1 that minimizes SSE.

$$\frac{\partial SSE}{\partial \theta_0} = 0 \Rightarrow \sum_{i=1}^{n} \left[2[y_i - (\theta_0 + \theta_1 x_i)](-1) \right] = 0 \qquad \frac{\partial(-\theta_0)}{\partial \theta_0}$$

$$\Rightarrow \sum_{i=1}^{n} y_i - n\theta_0 - \theta_1 \sum_{i=1}^{n} x_i = 0 \quad \text{—— (1)}$$

$$\frac{\partial SSE}{\partial \theta_1} = 0 \Rightarrow \sum_{i=1}^{n} 2[y_i - (\theta_0 + \theta_1 x_i)](-x_i) = 0 \qquad \frac{\partial(-\theta_1 x_i)}{\partial \theta_1}$$

$$\sum_{i=1}^{n} (x_i y_i) - \theta_0 \sum_{i=1}^{n} x_i - \theta_1 \sum_{i=1}^{n} x_i^2 = 0 \quad \text{—— (2)}$$

By solving equations (1) and (2), we get

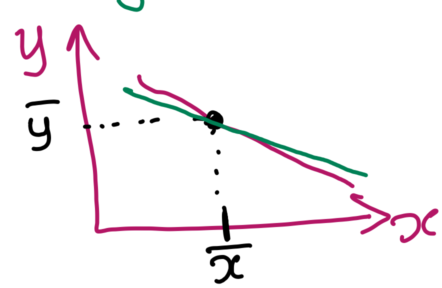
$$\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x} \quad \text{where } \bar{y} = \frac{\sum_1^n y_i}{n} \text{ and } \bar{x} = \frac{\sum_1^n x_i}{n}$$

$$\hat{\theta}_1 = \frac{n \sum_1^n (x_i y_i) - \left(\sum_1^n x_i\right)\left(\sum_1^n y_i\right)}{n \sum_1^n (x_i^2) - \left(\sum_1^n x_i\right)^2}$$

The fitted Least Squares simple linear regression model is

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$$

This line goes through $(\bar{x}, \bar{y})$ on scatterplot.



$$y_i = \theta_0 + \theta_1 x_i$$

$i=1;\quad y_1 = \theta_0 + \theta_1 x_1$

$i=2;\quad y_2 = \theta_0 + \theta_1 x_2$

$\vdots$

$i=n;\quad y_n = \theta_0 + \theta_1 x_n$

$$\Longrightarrow \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{n\times 2,\ X} \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}}_{2\times 1}$$

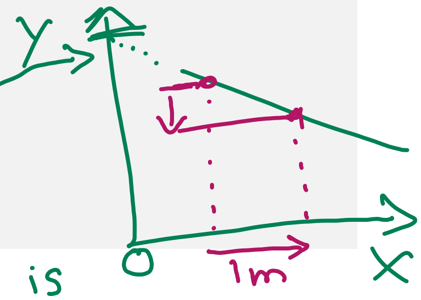
Design matrix

```
%%
% Fit and estimate a linear regression model
% The following function returns a vector of estimated coefficients %
% regress(y,x)
X =[ones(size(D(:,1))), D(:,1)] % Create a matrix with column of 1s followed by
other x variables %
Y= D(:,2) % Creates vector of y variable %
b = regress(Y,X);
b % intercept followed by other coefficients %

b =

   38.8040
   -0.4359
```

$\hat{\theta}_0 = 38.8040$, $\hat{\theta}_1 = -0.4359$



Estimated Least Squares Regression model is

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 X$$

$$\text{Signal strength} = 38.8040 - (0.4359\ \text{Distance})$$

Slope of -0.4359 says that average signal strength decreases by 0.4359 dB when distance to transmitter increases by 1 m.

Y-intercept of 38.804 says average signal strength is 38.804 dB when the distance to transmitter is zero.