1. Closed-form solution.

$$\theta = (x^T x)^{-1} x^T y$$

$$x = \underbrace{\left.\begin{bmatrix} — x^i — \\ — x^i — \\ — x^i — \end{bmatrix}\right\} \text{row}}_{\text{columns}} \; m \times n \qquad y = \begin{bmatrix} y^i \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} \qquad \theta = \left.\begin{bmatrix} \theta^1 \\ \vdots \\ \theta^n \end{bmatrix}\right\} \text{unknown}$$

Structure....

$$X = \begin{bmatrix} 1 & — x^i — & y^i \\ \vdots & — x^i — & \vdots \\ \vdots & — x^i — & \vdots \\ 1 & —— & y_i \end{bmatrix}$$

Where
1_s = constants
x^i = fetures
y_i = output

$$\begin{bmatrix} — x^i — \\ — x^i — \\ — x^i — \end{bmatrix} \times \begin{bmatrix} \theta^i \\ \vdots \\ \theta^n \end{bmatrix} = \left.\begin{bmatrix} y_i \\ \vdots \\ y_n \end{bmatrix}\right\} \hat{y}$$

$$\boxed{Cost = (h_\theta{}^i - y)^2}$$

$$\Bigg(\underbrace{\begin{bmatrix} \hat{y}_i \\ \vdots \\ \hat{y}_i \end{bmatrix} - \begin{bmatrix} y_i \\ \vdots \\ y_i \end{bmatrix}}_{Z}\Bigg)^2$$

$$\boxed{Cost = Z^T \cdot Z}$$

$$\downarrow$$

$$\boxed{Cost = (\hat{y} - y)^T \cdot (\hat{y} - y)}$$

$$\boxed{Cost = (X\theta - y)^T \cdot (X\theta - y)}$$

$$\boxed{(AB)^T = B^T A^T}$$

Calculation

$$(x\theta - y)^T \cdot (x\theta - y)$$

$$(x^T\theta^T - y^T)\cdot(x\theta - y)$$

$$J(\theta) = x^T\theta^T \cdot x\theta - x^T\theta^T\cdot y - y^T\cdot x\theta + y^T\cdot y$$

Compute first derivative $\nabla_\theta J(\theta)$

$$\nabla_\theta J(\theta) = \nabla_\theta\left(x^T\theta^T\cdot x\theta - x^T\theta^T\cdot y - y^T\cdot x\theta + y^T\cdot y\right)$$

$$2x^Tx\theta - \nabla_\theta(\theta^Tx^Ty + \theta^Tx^Ty)$$

$$2x^Tx\theta - \nabla_\theta(2\theta^Tx^Ty)$$

$$2x^Tx\theta - 2x^Ty = 0$$

$$x^Tx\theta = x^Ty$$

$$\underbrace{(x^Tx)^{-1}(x^Tx)}_{1}\theta = (x^Tx)^{-1}x^Ty$$

$$1\ \theta = \boxed{(x^Tx)^{-1}x^Ty}$$ ⇒ Closed-form Solution...

Some derivative rules for matrix.

1. $\nabla_\theta a^T\theta = a$ is matrix
2. $\nabla_\theta \theta^T a = a$ is matrix
3. $\nabla_\theta \theta^T A\theta = 2A\theta$ where A is symmetric matrix

Gradient descent	Closed-form sol.
-> Iterative approach	-> Quick calculation
-> $\alpha = ?$ (step size) unknown	-> for smaller data-set.
-> for larger dataset	-> expensive for large dataset.
	-> No need of learning rate / step size