$$y^{(i)} = \theta^T x^{(i)}, \quad \theta = [\theta_0, \theta_1 \ldots]^T$$
$$x^{(i)} = [1, x_1^{(i)}, x_2^{(i)} \ldots]^T$$

$$y = [y^{(1)}, y^{(2)} \ldots y^{(N)}]^T, \quad N = \# \text{ DATA POINTS}, \quad M = \# \text{ FEATURES}$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_M^{(1)} \\ 1 & x_1^{(2)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(N)} & \cdots & x_M^{(N)} \end{bmatrix}$$

$\hat{\theta}$

$$y = X\theta$$

$$\nabla \|y - X\theta\|_2^2 =$$
$$\nabla_\theta [(y - X\theta)^T (y - X\theta)]$$
$$= \nabla_\theta (\cancel{y^T y} - 2y^T X\theta + \theta^T X^T X\theta) = 0$$

0

$$\boxed{-2y^T X + 2X^T X\theta} = 0$$

$$X^T y = (X^T X)\theta \rightarrow \boxed{\theta = \boxed{(X^T X)^{-1} X^T} y}$$

$r_i = y_i - X_i\theta$ STACK ROWS $\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$

$r = y - X\theta$

$\theta^+ = \theta - (J_r^T J_r)^{-1} J_r^T r(\theta)$

$J_r = \frac{\partial}{\partial\theta}[r] = \frac{\partial}{\partial\theta}[y - X\theta]$

$= \begin{bmatrix} \frac{\partial}{\partial\theta} r_1 \\ \vdots \\ \partial/\partial\theta\, r_n \end{bmatrix} = \boxed{-X}$

$y = X\theta + \epsilon$, ϵ IS ZERO MEAN NOISE

θ ARE DETERMINISTIC

$\boxed{\min_\theta \{\|y - X\theta\|_2^2\}}$