Jan 20: Least squares derivation

for a function $\hat{y} = \theta_0 + \theta_1 x$,

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x_i - y_i)^2 = \frac{1}{m} \sum g(\theta_0, \theta_1)^2$$

↑ chain rule

$$\frac{\partial MSE}{\partial \theta_0} = \frac{2}{m} \sum_i (\theta_0 + \theta_1 x_i - y_i) \cdot 1$$

$$\frac{\partial MSE}{\partial \theta_1} = \frac{2}{m} \sum_i (\theta_0 + \theta_1 x_i - y_i) \cdot x_i$$

$$\sum_i (\theta_0 + \theta_1 x_i - y_i) = 0 \quad (1), \qquad \sum_i (\theta_0 + \theta_1 x_i - y_i) x_i = 0 \quad (2)$$

$$= \underbrace{\sum_i \theta_0}_{m\theta_0} + \sum_i \theta_1 x_i - \sum_i y_i \rightarrow \theta_0 = \frac{\sum_i y_i - \theta_1 \sum_i x_i}{m}$$

$$= \mu_y - \theta_1 \mu_x \quad //$$

(2) $$\sum_i \theta_0 x_i + \sum_i \theta_1 x_i^2 - \sum_i y_i x_i = 0$$

$$= \theta_0 \sum_i x_i + \theta_1 \sum_i x_i^2 - \sum_i y_i x_i = 0$$

$$= (\mu_y - \theta_1 \mu_x) \sum_i x_i + \ldots \text{etc}$$

$$= \mu_y \sum_i x_i - \theta_1 \left(\mu_x \sum_i x_i - \sum_i x_i^2\right)$$

$$\theta_1 = \frac{\mu_y \sum_i x_i - \sum_i y_i x_i}{\mu_x \sum_i x_i - \sum_i x_i^2} \quad //$$

Matrix form

$$\nabla_\theta \left(\frac{1}{m} (X\vec{\theta} - \vec{y})^T (X\vec{\theta} - \vec{y}) \right) = 0$$

$AB^T = B^T A^T$

$$= \nabla_\theta \frac{1}{m} (\vec{\theta}^T X^T - \vec{y}^T)(X\vec{\theta} - \vec{y})$$

$$= \nabla_\theta \frac{1}{m} \left(\vec{\theta}^T X^T X \vec{\theta} - \underbrace{\vec{\theta}^T X^T \vec{y} - \vec{y} X \vec{\theta}}_{\text{vectors, order doesn't matter}} + \vec{y}^T \vec{y}\right)$$

$$= \nabla_\theta \frac{1}{m} \left(\vec{\theta}^T X^T X \vec{\theta} - 2\theta^T X^T \vec{y} + \vec{y}^T \vec{y}\right)$$

$\vec{\theta}^T \vec{\theta} = \sum_i \theta_i^2 \rightarrow \frac{\partial}{\partial \theta} = 2\vec{\theta}^T \vec{\theta}$

$$\nabla_\theta = \frac{1}{m} (2X^T X \vec{\theta} - 2X^T \vec{y}) = \frac{2X^T}{m} (X\vec{\theta} - \vec{y}) \;//$$

The gradient to descend!

closed form solution: $\nabla_\theta = 0$

$$\vec{\theta} = (X^T X)^{-1} (X^T \vec{y}) \;//$$