1 Model

1.1 Model

$$\hat{y} = ax + b = \theta^T \hat{x}$$

where $\theta = \begin{bmatrix} a \\ b \end{bmatrix}, \hat{x} \begin{bmatrix} x \\ 1 \end{bmatrix}$.

1.2 Cost Function

$$\mathcal{L}(\theta) = \frac{1}{2} \sum_{i=1}^{N} \left(y^{(i)} - \hat{y}^{(i)} \right)^2 = \frac{1}{2} \sum_{i=1}^{N} \left(y^{(i)} - \theta^T \hat{x} \right)^2$$

To optimize $\mathcal{L}(\theta)$, calculate $\frac{\partial \mathcal{L}}{\partial \theta} = 0$.

2 Gradient

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}^n$. Then if $y = f(x)$, the gradient is defined as

$$\frac{\partial \mathcal{L}}{\partial \theta} = \nabla_x y = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$$