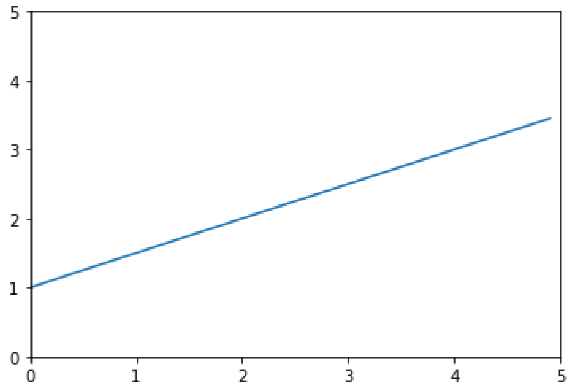
Linear Regression with One Variable

- Model representation:

$$h_\theta(x) = \theta_0 + \theta_1 x$$



- Cost function intuition:

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{1}{2m}\sum_{i=1}^{m}(\begin{pmatrix} 1, x^{(1)} \\ 1, x^{(2)} \\ \dots \\ 1, x^{(m)} \end{pmatrix}\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} - \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(m)} \end{pmatrix})^2$$

Goal:

$$minimize_{\theta_0, \theta_1} \quad J(\theta_0, \theta_1)$$

- Gradient descent:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta\left(x^{(i)}\right) - y^{(i)})^2$$

$$= \frac{1}{2m}\sum_{i=1}^{m}(\left(\theta_0 + \theta_1 x^{(i)}\right) - y^{(i)})^2$$

$$= \frac{1}{2m}(\left(\left(\theta_0 + \theta_1 x^{(1)}\right) - y^{(1)}\right)^2 + \left(\left(\theta_0 + \theta_1 x^{(2)}\right) - y^{(2)}\right)^2 + \cdots$$
$$+ (\left(\theta_0 + \theta_1 x^{(m)}\right) - y^{(m)})^2)$$

$$\to \frac{\partial J}{\partial \theta_0} = \frac{1}{2m}(2\left(\left(\theta_0 + \theta_1 x^{(1)}\right) - y^{(1)}\right) + 2\left(\left(\theta_0 + \theta_1 x^{(2)}\right) - y^{(2)}\right) + \cdots$$
$$+ 2(\left(\theta_0 + \theta_1 x^{(m)}\right) - y^{(m)}))$$

$$= \frac{1}{m}(\left(\left(\theta_0 + \theta_1 x^{(1)}\right) - y^{(1)}\right) + \left(\left(\theta_0 + \theta_1 x^{(2)}\right) - y^{(2)}\right) + \cdots + (\left(\theta_0 + \theta_1 x^{(m)}\right)$$
$$- y^{(m)}))$$

$$= \frac{1}{m}(\left(h_\theta\left(x^{(1)}\right) - y^{(1)}\right) + \left(h_\theta\left(x^{(2)}\right) - y^{(2)}\right) + \cdots + (h_\theta\left(x^{(m)}\right) - y^{(m)}))$$

$$= \frac{1}{m}\sum_{i=1}^{m}(h_\theta\left(x^{(i)}\right) - y^{(i)})$$

$$= \frac{1}{m}\sum_{i=1}^{m}(\begin{pmatrix} 1, x_1 \\ 1, x_2 \\ \cdots \\ 1, x_m \end{pmatrix}_{m*2} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}_{2*1} - \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{pmatrix})$$

$$\rightarrow \frac{\partial J}{\partial \theta_1} = \frac{1}{2m}(2\left(\left(\theta_0 + \theta_1 x^{(1)}\right) - y^{(1)}\right)x^{(1)} + 2\left(\left(\theta_0 + \theta_1 x^{(2)}\right) - y^{(2)}\right)x^{(2)} + \cdots$$
$$+ 2(\left(\theta_0 + \theta_1 x^{(m)}\right) - y^{(m)})x^{(m)})$$

$$= \frac{1}{m}((h_\theta\left(x^{(1)}\right) - y^{(1)})x^{(1)} + \left(h_\theta\left(x^{(2)}\right) - y^{(2)}\right)x^{(2)} + \cdots + (h_\theta\left(x^{(m)}\right)$$
$$- y^{(m)})x^{(m)})$$

$$= \frac{1}{m}\sum_{i=1}^{m}(h_\theta\left(x^{(m)}\right) - y^{(m)})x^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m}(\begin{pmatrix} 1, x_1 \\ 1, x_2 \\ \cdots \\ 1, x_m \end{pmatrix}_{m*2} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}_{2*1} - \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{pmatrix}) \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{pmatrix}$$

Repeat until convergence:

$$\theta_j \coloneqq \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \;\; for\ j = 0\ and\ j = 1$$