$W = [W_0 \quad W_1 \quad W_2 \quad \cdots \quad W_n] \rightarrow$ randomly initialized

$$W_j := W_j - \alpha \frac{\partial}{\partial W_j} J(W)$$

$$J(W) = \frac{1}{2m} \sum_{i=1}^{m} \left(h(X_i) - Y_i \right)^2 \rightarrow \text{MSE}$$

m = number of training samples

n = number of features

$$h(X_i) = \sum_{j=0}^{n} W_j \, x_{i,j}$$

$$X_i = [x_{i,1} \quad x_{i,2} \quad \cdots \quad x_{i,n}]$$

$$\frac{\partial}{\partial w_j} J(w) = \frac{\partial}{\partial w_j}\left[\frac{1}{2m}\sum_{i=1}^{m}\left(h(x_i) - y_i\right)^2\right]$$

$$= \frac{1}{2m}\sum_{i=1}^{m} 2\left(h(x_i) - y_i\right)\frac{\partial}{\partial w_j}\left(h(x_i) - y_i\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h(x_i) - y_i\right)\frac{\partial}{\partial w_j}\left(\sum_{j=0}^{n} w_j x_{i,j} - y_i\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(h(x_i) - y_i\right) x_{i,j}$$

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(w)$$

$$= w_j - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(h(x_i) - y_i\right) x_{i,j}$$