Gradient Descent

Gradient descent, or back propagation, is a first order iterative optimization algorithm used to minimize the cost function when training artificial neural networks.

$$
\begin{aligned}
J(\theta_0, \theta_1) &= \frac{1}{2m} \sum_{i=1}^{m} [h_\theta(x_i) - y_i]^2 \\
\theta_j &= \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \\
\frac{\partial}{\partial \theta} J_\theta &= \frac{\partial}{\partial \theta} \frac{1}{2m} \sum_{i=1}^{m} [h_\theta(x_i) - y_i]^2 \\
&= \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i) \frac{\partial}{\partial \theta} (\theta x_i - y) \\
&= \frac{1}{m} \sum_{i=1}^{m} [(h_\theta(x_i) - y) x_i] \\
\theta_j &:= \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} [(h_\theta(x_i) - y_i) x_i]
\end{aligned}
$$