Gradient Descent

Batch Gradient Descent

Initialize θ

Repeat {

$$\theta_j \leftarrow \theta_j - \alpha \frac{1}{n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}} \left(\mathbf{x}_i \right) - y_i \right) x_{ij} \qquad \text{for } j = 0...d$$

}

$$\frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta})$$

Stochastic Gradient Descent

Initialize θ

Randomly shuffle dataset

Repeat { (Typically 1 − 10x)

For $i = 1...n$, **do**

$$\theta_j \leftarrow \theta_j - \alpha \left(h_{\boldsymbol{\theta}} \left(\mathbf{x}_i \right) - y_i \right) x_{ij} \qquad \text{for } j = 0...d$$

}

$$\frac{\partial}{\partial \theta_j} \text{cost}_{\boldsymbol{\theta}} (\mathbf{x}_i, y_i)$$