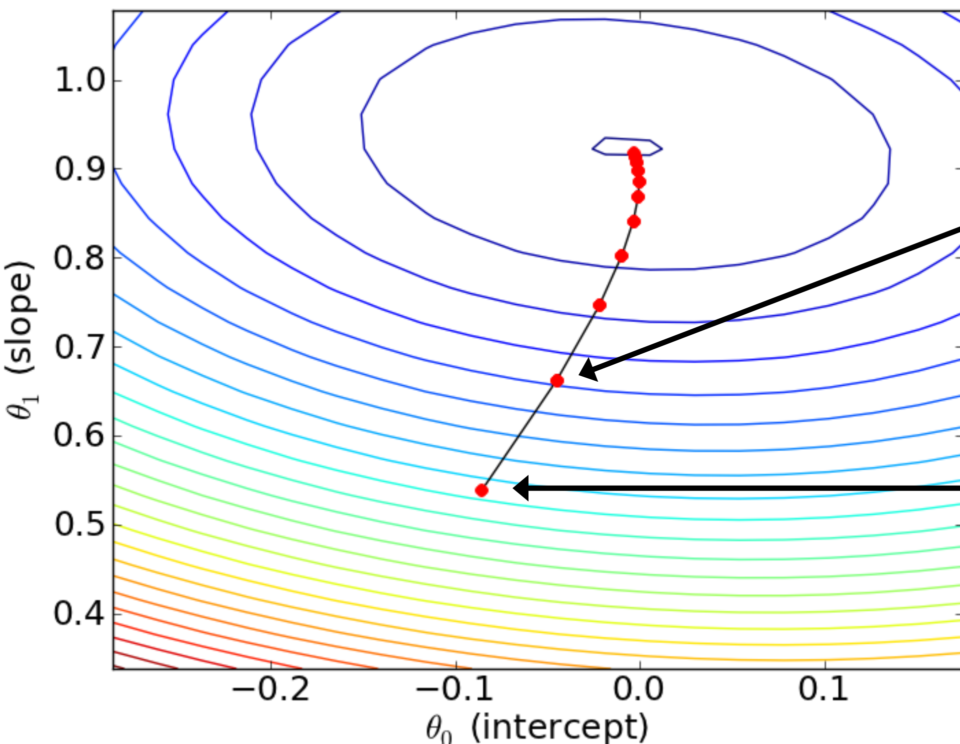
$$\theta_0^{(i+1)} \leftarrow \theta_0^{(i)} + \alpha \frac{\delta}{\delta\theta_0} J(\theta^{(i)})$$

$$\theta_1^{(i+1)} \leftarrow \theta_1^{(i)} + \alpha \frac{\delta}{\delta\theta_1} J(\theta^{(i)})$$



$$(\theta_0^{(2)}, \theta_1^{(2)})$$

$$(\theta_0^{(1)}, \theta_1^{(1)})$$

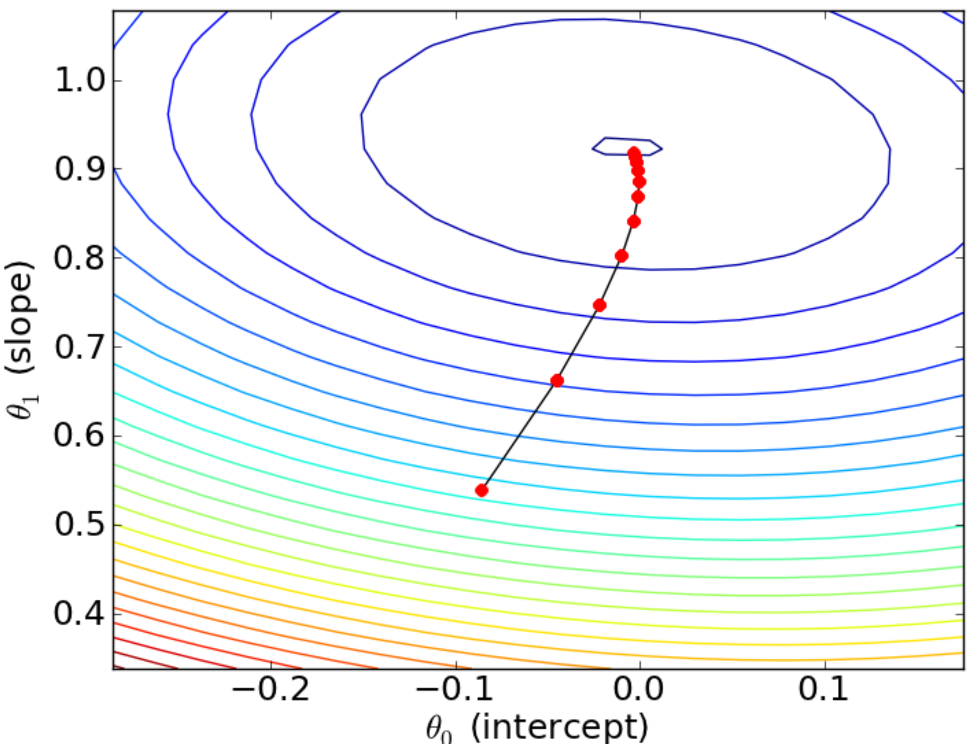
$$\theta_0^{(i+1)} \leftarrow \theta_0^{(i)} + \alpha \frac{\delta}{\delta\theta_0} J(\theta^{(i)})$$

$$\theta_1^{(i+1)} \leftarrow \theta_1^{(i)} + \alpha \frac{\delta}{\delta\theta_1} J(\theta^{(i)})$$

Learning Rate

Partial derivative with respect to θ1

Cost Function



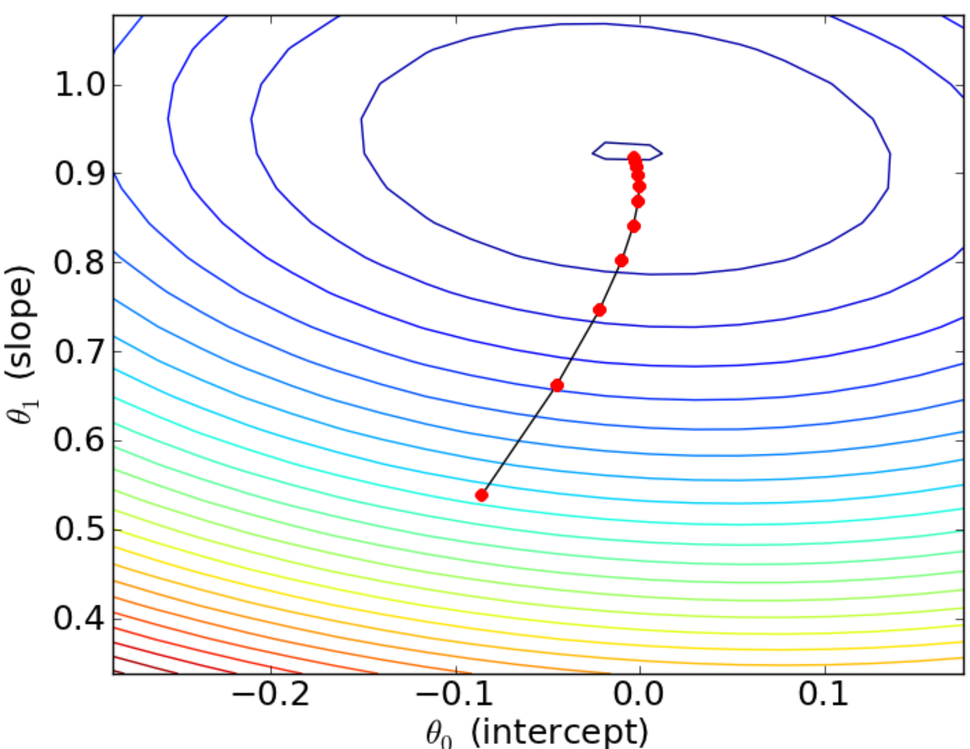
Model at *i*

$$\theta_0^{(i+1)} \leftarrow \theta_0^{(i)} + \alpha \frac{\delta}{\delta\theta_0} \underbrace{\frac{1}{2} \sum_{k=1}^{n} \left(h^{(i)}(x_k) - y_k\right)^2}_{\text{Cost Function}}$$

Learning Rate

Gradient

Cost Function



$$\frac{\delta}{\delta\theta_1}\frac{1}{2}\sum_{k=1}^{n}(h(x_k)-y_k)^2$$

$$\sum_{k=1}^{n}2\frac{1}{2}(h(x_k)-y_k)\frac{\delta}{\delta\theta_1}(h(x_k)-y_k)$$

$$\sum_{k=1}^{n}(h(x_k)-y_k)\frac{\delta}{\delta\theta_1}(\theta_0+\theta_1 x_k-y_k)$$

$$\sum_{k=1}^{n}(h(x_k)-y_k)x_k$$

$$\theta_1^{(i+1)} \leftarrow \theta_1^{(i)} + \alpha \frac{\delta}{\delta\theta_0} \frac{1}{2} \sum_{k=1}^{n} (h^{(i)}(x_k) - y_k)^2$$

$$\theta_1^{(i+1)} \leftarrow \theta_1^{(i)} + \alpha \sum_{k=1}^{n} (\theta_0^{(i)} + \theta_1^{(i)}(x_k) - y_k) x_k$$

while not converged:

$$\theta_0^{(i+1)} \leftarrow \theta_0^{(i)} + \alpha \sum_{k=1}^{n} (\theta_0^{(i)} + \theta_1^{(i)}(x_k) - y_k) * 1.0$$

$$\theta_1^{(i+1)} \leftarrow \theta_1^{(i)} + \alpha \sum_{k=1}^{n} (\theta_0^{(i)} + \theta_1^{(i)}(x_k) - y_k) x_k$$

Some questions

- Initialization
  - Where do you drop the ball?
  - “small random values”
- Step size
  - We don’t really “roll,” we “jump” in the direction of steepest descent
  - How far should we jump?
  - Too far and you might hop over the minimum and raise the function vaue
  - Too small and performance is bad