线性回归梯度下降　　　迭代求 $L(\theta)$ 最小值时模型参数 θ

这里 $L(\theta)$ 用均方误差（MSE）

单特征：　　$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$　　　学习率：α

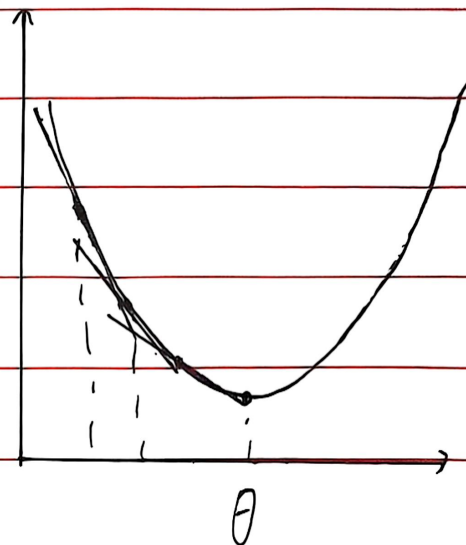$$L(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

$L(\theta)$ 是类二次函数，随机初始化 θ_0, θ_1

$$\theta_1 = \theta_1 - \alpha \frac{\partial L(\theta)}{\partial \theta_1}$$

$$\theta_0 = \theta_0 - \alpha \frac{\partial L(\theta)}{\partial \theta_0}$$

迭代更新参数，像下楼梯一步步走到

最小值

α 类似步长，过长则走到另一边，可能离 min 越来越远

过小则迭代慢

多特征：　　$L(\vec{\theta}) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\vec{\theta}}(X) - \vec{y})^2$

$$\vec{\theta} = \vec{\theta} - \alpha \frac{\partial L(\vec{\theta})}{\partial \vec{\theta}}$$

$$= \vec{\theta} - \frac{\alpha}{m} \cdot \sum_{i=1}^{m} (h_{\vec{\theta}}(X) - \vec{y})$$