Vazyme

用一阶泰勒展开式证明梯度下降：

$f(\theta)=f(\theta_0)+f'(\theta_0)\cdot(\theta-\theta_0)$

设θ_0是原点，θ是向下步后所处的点.

min：$f(\theta)-f(\theta_0)=(\theta-\theta_0)\cdot f'(\theta_0)$

用向量$\vec{v}$表示$\theta-\theta_0$：$\vec{v}=\theta-\theta_0$

$f(\theta)-f(\theta_0)=\vec{v}\cdot f'(\theta_0)=\|\vec{v}\|\cdot\|f'(\theta_0)\|\cdot\cos\alpha$.

要使$f(\theta)-f(\theta_0)$最小，$\cos\alpha=-1$，$\vec{v}$与$f'(\theta_0)$方向相反。

$\vec{v}$：$\vec{v}=-\eta\cdot f'(\theta_0)$

因为：$\vec{v}=\theta-\theta_0$，$\theta-\theta_0=-\eta\cdot f'(\theta_0)$

$\theta=\theta_0-\eta\cdot f'(\theta_0)$

因为要满足一阶泰勒展开式，所以

$\theta\to\theta_0$，$\eta\cdot f'(\theta_0)$要小，η要小，学习率要小.

Vazyme