

算法 8.1 随机梯度下降(SGD) 在第 k 个训练迭代的更新

Require: 学习速率 ϵ_k

Require: 初始参数 θ

while 没有达到停止准则 **do**

从训练集中采包含 m 个样本 $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}\}$ 的minibatch, 对应目标为 $\mathbf{y}^{(i)}$ 。

计算梯度估计: $\hat{\mathbf{g}} \leftarrow +\frac{1}{m} \nabla_{\theta} \sum_i L(f(\mathbf{x}^{(i)}; \theta), \mathbf{y}^{(i)})$

应用更新: $\theta \leftarrow \theta - \epsilon \hat{\mathbf{g}}$

end while



