Require: 学习速率 ϵ_k Require: 初始参数 $\boldsymbol{\theta}$ while 没有达到停止准则 do 从训练集中采包含 m 个样本 $\{\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(m)}\}$ 的minibatch,对应目标为 $\boldsymbol{y}^{(i)}$ 。 计算梯度估计: $\boldsymbol{g} \leftarrow +\frac{1}{m}\nabla_{\boldsymbol{\theta}}\sum_{i}L(f(\boldsymbol{x}^{(i)};\boldsymbol{\theta}),\boldsymbol{y}^{(i)})$ 应用更新: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \hat{\boldsymbol{g}}$ end while	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		
while 没有达到停止准则 do 从训练集中采包含 m 个样本 $\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(m)}\}$ 的minibatch,对应目标为 $\boldsymbol{y}^{(i)}$ 。 计算梯度估计: $\hat{\boldsymbol{g}}\leftarrow+\frac{1}{m}\nabla_{\boldsymbol{\theta}}\sum_{i}L(f(\boldsymbol{x}^{(i)};\boldsymbol{\theta}),\boldsymbol{y}^{(i)})$ 应用更新: $\boldsymbol{\theta}\leftarrow\boldsymbol{\theta}-\epsilon\hat{\boldsymbol{g}}$			
从训练集中采包含 m 个样本 $\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(m)}\}$ 的minibatch,对应目标为 $\boldsymbol{y}^{(i)}$ 。 计算梯度估计: $\hat{\boldsymbol{g}}\leftarrow+\frac{1}{m}\nabla_{\boldsymbol{\theta}}\sum_{i}L(f(\boldsymbol{x}^{(i)};\boldsymbol{\theta}),\boldsymbol{y}^{(i)})$ 应用更新: $\boldsymbol{\theta}\leftarrow\boldsymbol{\theta}-\epsilon\hat{\boldsymbol{g}}$	Require: 初始参数 $ heta$		
计算梯度估计: $\hat{\boldsymbol{g}} \leftarrow +\frac{1}{m} \nabla_{\boldsymbol{\theta}} \sum_{i} L(f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}), \boldsymbol{y}^{(i)})$ 应用更新: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \hat{\boldsymbol{g}}$	while 没有达到停止准则 do		
计算梯度估计: $\hat{\boldsymbol{g}} \leftarrow +\frac{1}{m} \nabla_{\boldsymbol{\theta}} \sum_{i} L(f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}), \boldsymbol{y}^{(i)})$ 应用更新: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \hat{\boldsymbol{g}}$	从训练集中采包含	m 个样本 $\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(m)}\}$ 的minibatch,对应目标为 $\boldsymbol{y}^{(i)}$ 。	
应用更新: $oldsymbol{ heta} \leftarrow oldsymbol{ heta} - \epsilon \hat{oldsymbol{g}}$	计算梯度估计: \hat{g} \leftarrow	$-+rac{1}{m} abla_{m{ heta}}\sum_{i}L(f(m{x}^{(i)};m{ heta}),m{y}^{(i)})$	
end while	应用更新: $\theta \leftarrow \theta$	$-\epsilon\hat{m{g}}$	
	end while		



