1. 准备训练数据集 D = {(x,1, x,2, ..., x,n, y)} // x\_i ∈ ℝ^{n+1} (含偏置项)

2. 随机初始化权重 w\_0, w\_1, ..., w\_n

设置学习率 α，批量大小 b

3. while 未收敛 do:

4. 在D中随机选择 b 个样本，记为小批量 D^b

5. 初始化梯度累积值：∀j, ∂J/∂w\_j ← 0

6.

7. for 每个样本 (x\_i, y\_i) in D^b do:

8. 计算预测值 h(x\_i) = Σ\_{k=0}^n w\_k·x\_i^{(k)} // 前向传播

9. 计算误差项 δ = h(x\_i) - y\_i

10.

11. for j = 0 to n do: // 反向传播

12. ∂J/∂w\_j += δ \* x\_i^{(j)} // 累积样本梯度

13. end for

14. end for

15.

16. for j = 0 to n do: // 参数更新

17. w\_j ← w\_j - α \* (∂J/∂w\_j / b) // 取平均梯度

18. end for

19. end while