

实验报告

李鹏宇

1 作业任务

在本次作业中，我实现了以下功能：

- 加载 MNIST 数据集为 numpy array
- 定义一个两层的神经网络结构，并实现 forward 和 softmax_loss 计算
- 实现 SGD 和 Adam 优化器。

2 优化器介绍

2.1 SGD

在本代码中，SGD 实现的是 mini-batch 梯度下降，它在梯度的准确性和运算速度上做了权衡。在每次更新中，SGD 使用了小批量数据进行计算，以提高计算效率，并且在一定程度上减小了噪声。

2.2 Adam 优化器

Adam 通过计算梯度的一阶矩，使得梯度的更新更加平滑、稳定。通过计算梯度的二阶矩，当梯度的二阶矩估计很大时，参数更新逐步减小，从而避免了震荡；当该值很小时，参数更新会变大，帮助跳出梯度平坦的 plateau。

同时，由于梯度矩估计在最开始偏向于零，所以 Adam 使用偏差矫正技术来保留大部分原始的梯度，随着时间推移逐步减少偏差矫正的程度。

3 实验结果

3.1 SGD

参数设置：hidden_dim = 100, epochs = 20, lr = 0.5, batch = 100。

Epoch	Train Loss	Train Err	Test Loss	Test Err
0	0.16646	0.05183	0.17233	0.05360
1	0.11469	0.03587	0.12737	0.03950
2	0.08929	0.02797	0.10957	0.03450
3	0.07162	0.02227	0.09843	0.03060
4	0.05676	0.01725	0.08784	0.02650
5	0.04666	0.01387	0.08137	0.02340
6	0.04021	0.01158	0.07793	0.02200
7	0.03547	0.01018	0.07561	0.02190
8	0.03177	0.00897	0.07439	0.02160
9	0.02881	0.00800	0.07320	0.02110
10	0.02645	0.00718	0.07194	0.02050
11	0.02440	0.00617	0.07120	0.02060
12	0.02279	0.00540	0.07050	0.02050
13	0.02136	0.00493	0.06992	0.02020
14	0.02019	0.00450	0.06936	0.02050
15	0.01934	0.00405	0.06895	0.02050
16	0.01873	0.00372	0.06880	0.02030
17	0.01830	0.00352	0.06867	0.02030
18	0.01805	0.00332	0.06860	0.02020

Epoch	Train Loss	Train Err	Test Loss	Test Err
19	0.01790	0.00330	0.06858	0.02010

3.2 Adam

参数设置: hidden_dim = 100, epochs = 20, lr = 0.01, batch = 100, beta1 = 0.9, beta2 = 0.999。

Epoch	Train Loss	Train Err	Test Loss	Test Err
0	0.12610	0.03820	0.14098	0.04170
1	0.14850	0.04350	0.20755	0.05380
2	0.21262	0.05280	0.30831	0.06770
3	0.12772	0.03543	0.20081	0.04670
4	0.08930	0.02423	0.18767	0.03880
5	0.06453	0.01847	0.15823	0.03230
6	0.04210	0.01253	0.14928	0.02900
7	0.03450	0.01028	0.15454	0.02740
8	0.04560	0.01300	0.16871	0.03170
9	0.02874	0.00897	0.16800	0.02830
10	0.02208	0.00650	0.16753	0.02770
11	0.01333	0.00420	0.16268	0.02620
12	0.00995	0.00342	0.16749	0.02620
13	0.00713	0.00220	0.17184	0.02580
14	0.00449	0.00137	0.17470	0.02490
15	0.00243	0.00068	0.17380	0.02350
16	0.00213	0.00048	0.17423	0.02370
17	0.00112	0.00010	0.17219	0.02290
18	0.00085	0.00007	0.17179	0.02240
19	0.00076	0.00007	0.17040	0.02210