

Assignment-part1

李鹏宇

October 8, 2024

1 实验结果

采用 SGD 优化器， $learning_rate = 0.001$ ， $momentum = 0.9$ ，训练 10 个 epoch 模型在全部测试集上的准确率为 60.05%。

Class	Accuracy (%)
0	61.20
1	70.00
2	51.60
3	44.50
4	54.90
5	50.80
6	66.10
7	63.70
8	74.10
9	63.60

Table 1: Accuracy of each class

2 损失曲线

See Figure1.

3 SGD optimizer

在 SGD 优化器中，momentum 参数是上一次训练梯度所乘的衰减因子，momentum 越大意味着更大程度地保持上一次训练的梯度方向。这使得当这一次训练梯度与上一次方向相近时，在相关的方向上加速收敛，相反时，有助于抑制振荡。总而言之，momentum 可以加快梯度下降的收敛速度，同时提高算法的稳定性，帮助模型在训练过程中跳出小的局部最小值，更接近全局最小值。

更新速度（velocity）的公式：

$$v_t = \beta v_{t-1} + \nabla J(\theta)$$

更新参数（parameters）的公式：

$$\theta_t = \theta_{t-1} - \eta v_t$$

其中： v_t 是在时间步 t 的速度。 β 是动量系数。 $\nabla J(\theta)$ 是目标函数 J 关于参数 θ 的梯度。 η 是学习率。 θ_t 是在时间步 t 的参数值。

从 figure 2 可以看到 loss 下降的总量逐渐小，当 momentum=0.1 时，loss 仅仅从 2.3 降至 2.05，收敛速度变慢，这说明 momentum 能加速训练过程。随着 momentum 减小，训练过程变得更加不稳定，训练曲线并非平滑地下降至收敛，出现了锯齿形震荡，这说明 momentum 能使得训练过程稳定。

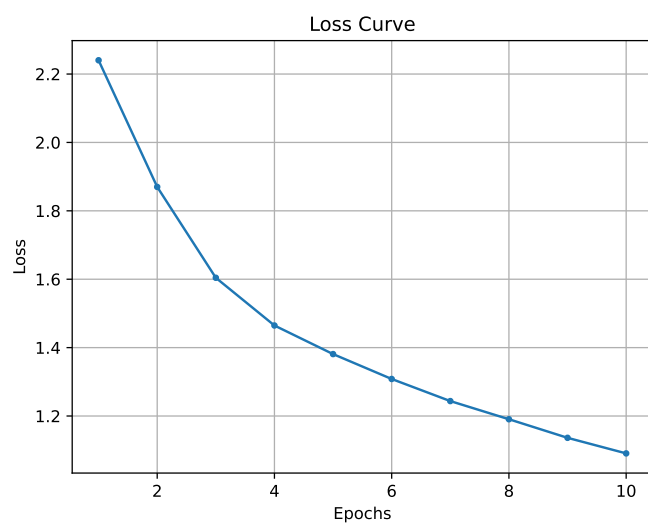
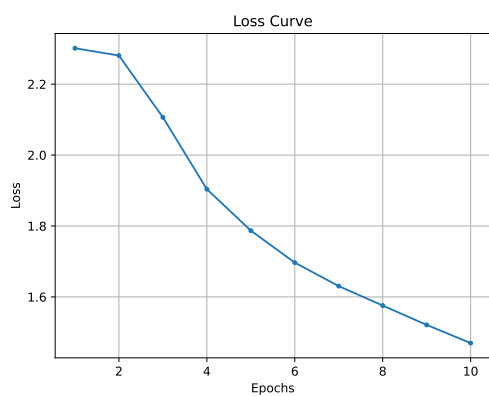
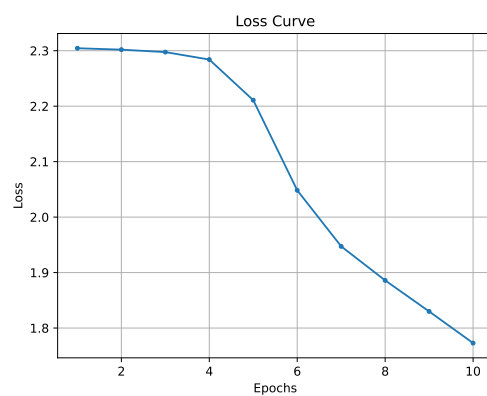


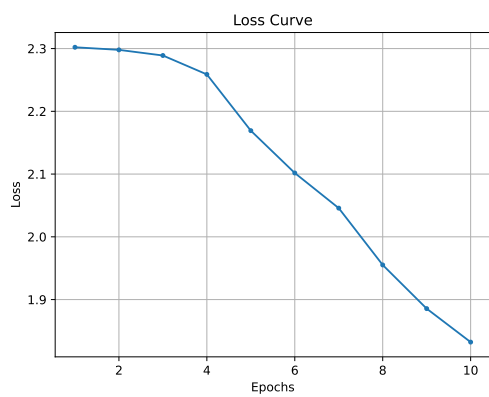
Figure 1: Loss curve for the training process



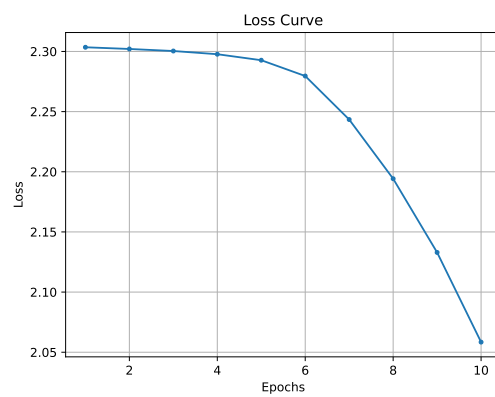
(a) momentum = 0.7



(b) momentum = 0.5



(c) momentum = 0.3



(d) momentum = 0.1

Figure 2: 对比 momentum 不同取值