

# Homework 5

林宇辰 2200013211

December 2023

## 1 Define Classifiers

### 1.1 Linear classifier

用一个 linear 层即可.

### 1.2 Full-connected neural network classifier

使用最简单的两层网络，在两个全连接层之间加一个 ReLu 层来引入非线性。

## 2 Define loss function

### 2.1 SVM loss function

依据公式正常实现即可。

$$svmloss = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + margin)$$

### 2.2 Cross-entropy loss function

按照定义实现即可。但是在计算 softmax 的时候，要先将数据在每个维度上减去最大值再进行计算，这是为了数值稳定性（防止出现 nan）。

## 3 Implement the training and testing function

### 3.1 Train

训练时，首先获得数据集，然后封装出一个接口。接着在每个 epoch 上进行预测和反向传播，更新参数值。全部训练完之后就保存到 checkpoint 中。注意应该把模型和数据移到显卡上，可以大幅加快训练进程。

### 3.2 Test

调用训练好的模型，测试表现。

### 4 AdamW vs SGD optimizer

如图 1所示，SGD 能够更快地训练模型，它的准确率比较高同时损失函数低。但是 Adam 优化器训练的曲线比较平稳。

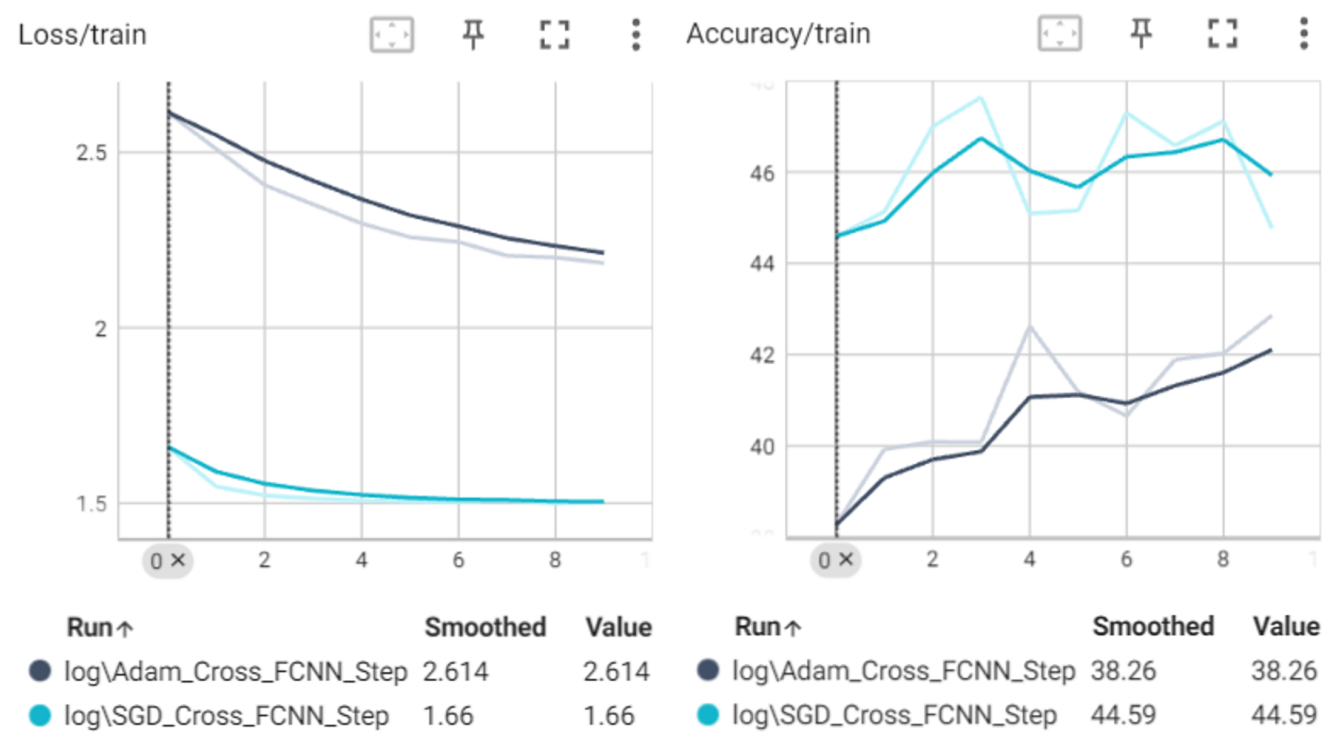


图 1: Adamw and SGD

### 5 SVM vs Corss-entropy loss

其实可以很显然地猜测，交叉熵的表现要比合页损失要好的多。支撑向量机在这种多维的复杂任务之下很难有很好的表现。如图 2所示，用交叉熵训练的模型表现要远好于用 svmloss 的模型。

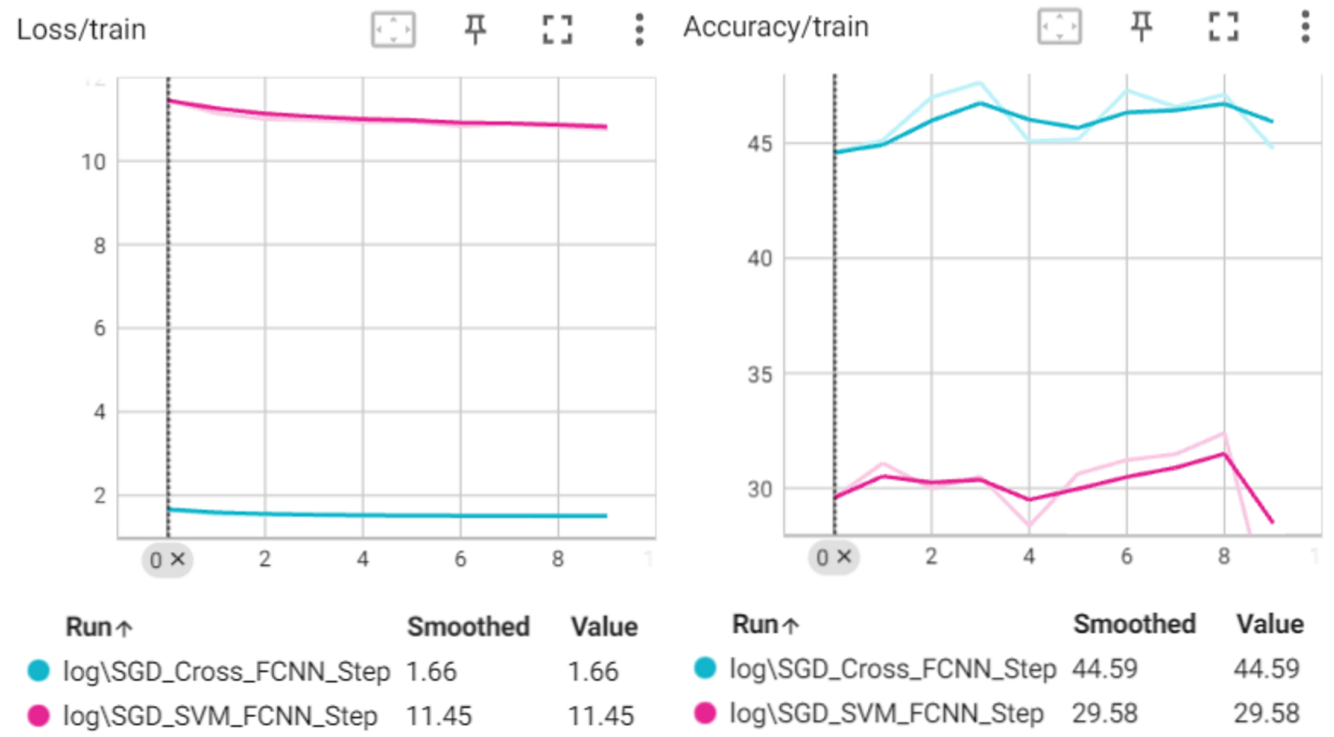


图 2: SVM and Corss-entropy

## 6 StepLR and CosineAnnealingLR scheduler

可以看出，cosine scheduler 要比 step scheduler 表现优秀很多，能更好地调整学习率，实现更高的准确率和更低的损失。

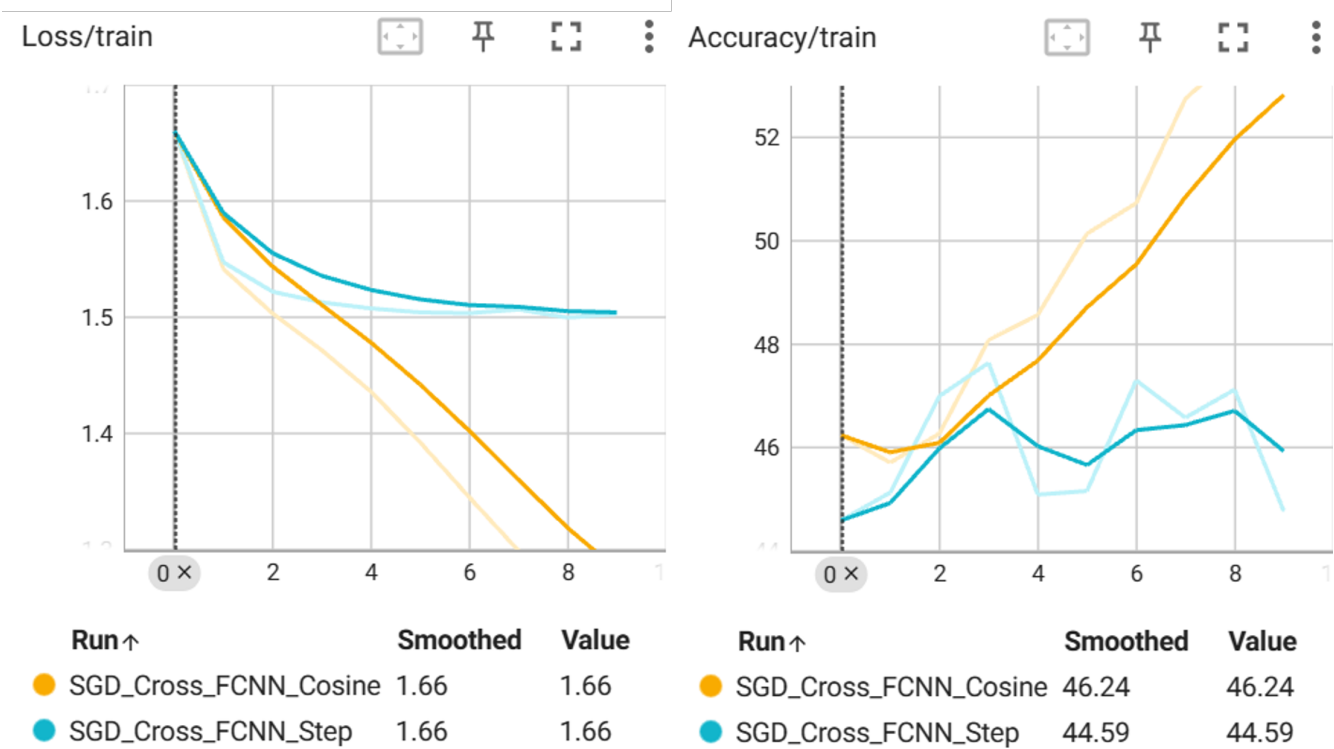


图 3: Step and CosineAnnealing

## 7 Final Setting

在使用含一个隐层，隐层维度为 1024 的全连接神经网络，损失函数为交叉熵损失函数，用 sgd 优化器和 cosine 调度器训练 10 个 epoch 的情况下，能够达到 54% 的准确率。在使用单一线性分类器，损失函数为交叉熵损失函数，用 sgd 优化器和 cosine 调度器训练 10 个 epoch 的情况下，能够达到 41% 的准确率。使用不同训练参数组合的情况如图 4所示。

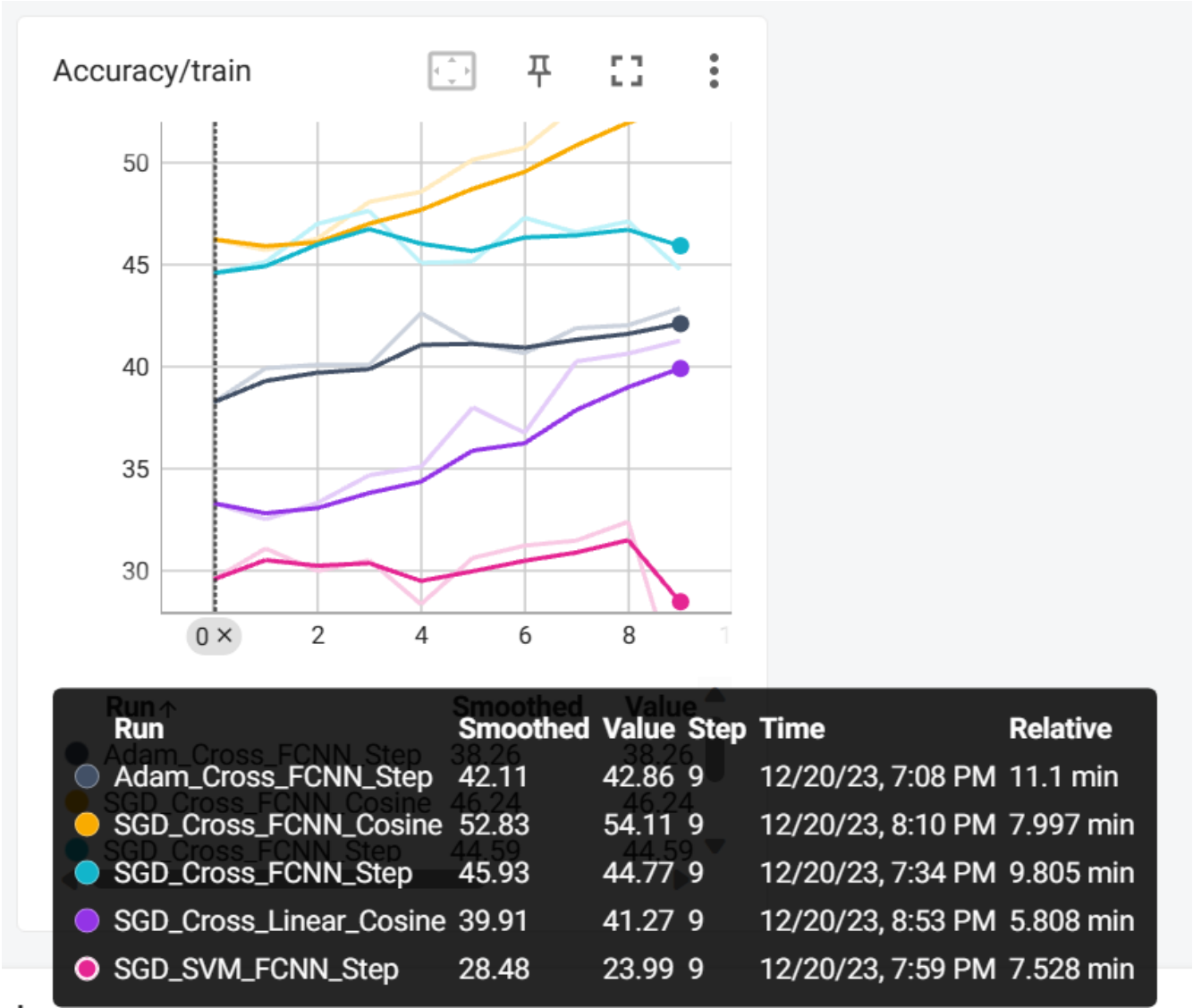


图 4: Final Setting