

報告 (Task 1.1 + Task 1.2)

黃茂誠 N26141826 2025/10/14

Task 1.1 : MLP

2.1 實作方式 (Implementation)

- Layers :
 - Linear : $y = XW + b$; backward : $\partial L / \partial W = X^\top \partial L / \partial y \cdot \partial L / \partial b = \sum \partial L / \partial y$
 - ReLU : $f(x) = \max(0, x)$; backward : mask 乘回傳
 - Sigmoid : $\sigma(x) = 1 / (1 + e^{-x})$; backward : $\sigma(x)(1 - \sigma(x))$
 - Softmax + CrossEntropy (合併做 numerically stable) :

$$\text{softmax}(z)_i = \frac{e^{z_i - \max(z)}}{\sum_j e^{z_j - \max(z)}}, \quad \mathcal{L} = -\frac{1}{N} \sum \log p(y_i)$$

backward : 對 logits 的梯度為 $(p - \text{onehot}(y)) / N$

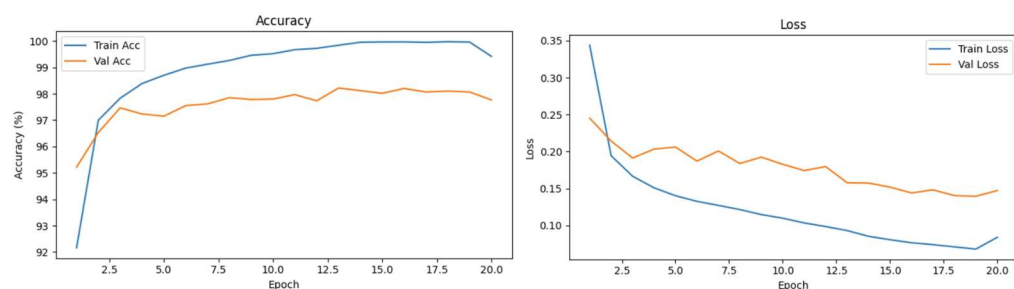
- Optimizer : SGD + momentum (可選 Nesterov), weight decay 作為 L2。
- Model : $784 \rightarrow 256 \rightarrow 128 \rightarrow 10$, activation 用 ReLU。
- Data : MNIST, Normalize(/255) (可選 z-score)。
- Split : train/valid=9:1 (或 K-Fold)。

Hyperparameters

Epochs	20
Batch size	128
Optimizer	SGD (lr=0.1, momentum=0.9, weight_decay=1e-4)
Hidden dims	[256, 128]
Activation	ReLU
Validation split	0.1

2.3 輸出與圖 (Results)

[TEST] loss=0.1302 acc=98.11%



Task 1.2 : ResNet-18

3.1 實作方式

Architecture :

ResNet-18 (BasicBlockx[2, 2, 2, 2])

CIFAR-friendly stem : 3x3, stride=1, padding=1, 無 MaxPool

接 AdaptiveAvgPool → FC(10)

Data Augmentation :

RandomCrop(32, padding=4) + RandomHorizontalFlip + ColorJitter。

Normalization : 使用 CIFAR-10 官方 mean/std

Optimizer / Scheduler (可比較多組) :

- SGD + CosineAnnealingLR (warmup)
- SGD + StepLR
- AdamW + Cosine

AMP : 可啟用 mixed precision。

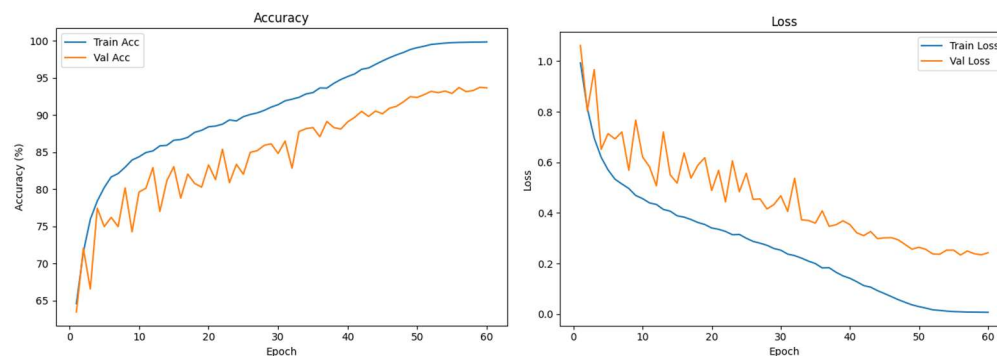
Model Summary : 總參數量、FLOPs

3.2 Hyperparameters

Epochs	60
Batch size	128
Optimizer	SGD (lr=0.1, momentum=0.9, wd=5e-4)
Scheduler	CosineAnnealingLR (T_max=epochs, warmup=5)
Augmentation	Crop + Flip + ColorJitter
Normalize	CIFAR10 mean/std

3.3 輸出與圖 (Results)

Test Loss: 0.2237 | Test Accuracy: 94.04%



遇到的問題與解決方式

4.1 Task 1.1 (MNIST from scratch)

- **Numerical stability**: softmax 前先減最大值避免 overflow; 交叉熵加 $1e-12$ 。
 - 解法: 實作 **SoftmaxCrossEntropy** 合併版, forward 用 shifted logits。
- **梯度爆炸/收斂慢**:
 - 解法: momentum=0.9、適度 **weight decay**, 學習率 0.1。
- **泛化不足**:
 - 解法: 增加 hidden units、加入 L2、早停或觀察 val 曲線調整 epoch。

4.2 Task 1.2 (ResNet-18, CIFAR-10)

- **未達 85%**:
 - 解法: 改用 **Cosine + warmup**、開 AMP、增強 ColorJitter。
- **下載資料**:
 - 先在可上網環境下載至 ./data/, 或共用快取。
- **顯存不足**:
 - 解法: 降 batch_size、開 AMP、pin_memory=True
- **訓練震盪**:
 - 解法: 提高 weight_decay 或小幅降低 LR; StepLR 改成 Cosine。