MLP&CNN for MNIST and CIFAR-10 (Score 10)

Homework 3 for Deep Learning, Autumn 2025

Deadline: 2025.11.10 12:00

Attention

- You need to submit all codes and a report (at least one page in PDF format).
- Plagiarism is not permitted.

1 Objective Questions (Score 3)

These questions are all multiple choice questions. One or more options may be correct.

1.1 (Score 0.5)

Which updating rules allow different learning rates for different parameters?

A. Stochastic gradient descent (SGD)

B. SGD+momentum

C. Adagrad

D. RMSProp

E. Adam

1.2 (Score 0.5)

Which updating rule has the early stopping problem?

A. Adagrad

B. RMSprop

C. Adam

1.3 (Score 0.5)

Which technique(s) can reduce overfitting?

A. Early stopping

B. Dropout

C. BatchNorm

1.4 (Score 0.5)

Is batchnorm layer usually placed before or after the nonlinear activation layer?

A. Before

B. After

1.5 (Score 0.5)

From GoogLeNet to ResNet, what’ s the key idea in model design?

A. Going deeper

B. Adding shortcuts

C. Using depthwise convolution

D. Using group convolution

1.6 (Score 0.5)

For a CNN, in general, most computational cost is on

A. Fully connected layers

B. Convolutional layers

C. Loss layer

D. BN layers

2 Programming Practice (Score 7)

In this homework, you are required to redo the MNIST classification with PyTorch or MindSpore framework. Additionally, you need to conduct the same classification experiment on the CIFAR-10 dataset .

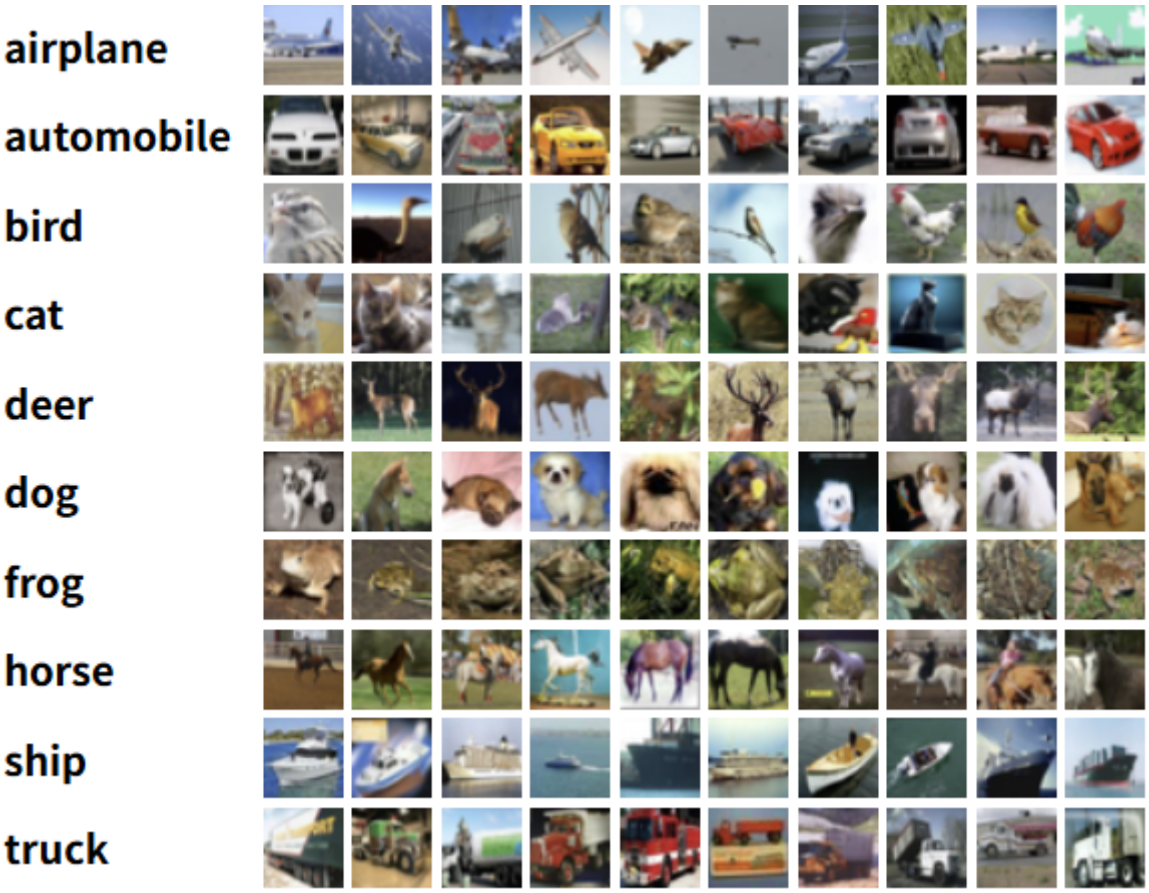
No starting codes. It all depends on you! You need to submit all codes and a short report with the following requirements:

2.1 DATASETS

MNIST

MNIST digits dataset is a widely used dataset for image classification in machine learning field. It contains 60,000 training examples and 10,000 testing examples. The digits have been size-normalized and centered in a fixed-size image. Each example is a 784×1 matrix, which is transformed from an original 28×28 grayscale image. Digits in MNIST range from 0 to 9. Some examples are shown below. Note: During training, information about testing examples should never be used in any form.

CIFAR-10



2.2 MLP&CNN for MNIST and CIFAR-10

(1) Basic Model Implementation

Implement two types of models:

1. MLP Model (Two-Layer Fully Connected)

Structure: Input (flattened image, e.g., MNIST 28x28=784) → Sigmoid → 256 (hidden layer) → Sigmoid → 10 (output, 10-class classification)

2. Pure Convolutional CNN Model (No Linear Layers)

Structure:

1. 1st Conv: Conv2d(1→16, kernel=3, stride=1, padding=1) → Sigmoid → MaxPool2d(kernel=2, stride=2)
2. 2nd Conv: Conv2d(16→32, kernel=3, stride=1, padding=1) → Sigmoid → MaxPool2d(kernel=2, stride=2)
3. Classification Conv: Conv2d(32→10, kernel=7, stride=1, padding=0) → Global Average Pooling (GAP) → Flatten to (batch, 10)

(2) Baseline Experiments (Score 3)

Follow the baseline config：

- Optimizer: SGD; Learning rate: 0.05; BatchSize: 64; Epochs: 15; Loss function: CrossEntropyLoss

And record loss/accuracy and plot curves for all experiments

(3) More Experiments (Score 4)

1. Switch the activation function from Sigmoid to ReLU. What are the differences compared to the previous one?

2. Replace the optimizer from SGD to Adam, observe the experimental results. Why are such results obtained? How to optimize it?

3. Add BatchNorm to observe its impact on experimental results, and plot the distribution of the output values of the second Sigmoid function.

4. Compare the parameter count, training performance, and convergence speed of MLP and CNN, and plot the differences.