

Handwritten digit recognition

Ziqiang LI 李子强(11510352)

Department of Computer Science and Engineering,
Southern University of Science and Technology
Email: 11510352@mail.sustc.edu.cn

1. Preliminaries

Handwritten digit recognition is a simplified problem of handwriting recognition problem. The model only need to classify 10 different digits. The training set, MNIST [2] was given by LeCun et al. 1998. I use a subset of it. In the implement, I use back propagation neural network (BP) with 2 hidden layers. Hidden layer 1 and 2 are activated by rectified linear unit and output layer is activated by softmax function.

2. Methodology

2.1. Required packages

- **Pickle:** reader for formatted dataset
- **numpy:** matrix calculation
- **matplotlib:** drawing the output figures

2.2. Main functions

- **Xavier initialization [1]**

$$W \sim U\left[-\sqrt{\frac{6}{n_j + n_{j+1}}}, \sqrt{\frac{6}{n_j + n_{j+1}}}\right]$$

W obeys the above uniform distribution.

- **Rectified Linear Unit (RELU):**
For hidden layer 1 and layer 2

$$f(x) = \max(0, x)$$

- **Softmax:** For output layer

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

- **cross entropy loss function:**

$$E(w, b) = -\sum_{k=1}^C t_k \ln y_k(x, w, b)$$

C is the dimension of t_k . t_k is a one-hot vector which is corresponding to the label of y , and $y_k(x, w)$ gives the output vector y by input vector x , weight matrix w and bias matrix b .

- **cross entropy loss function** (when back propagation):

$$E_{total}(w, b) = \frac{1}{N} \sum_{k=1}^N E(w^k, b^k) + \frac{\lambda}{2} \|w\|^2$$

$\frac{\lambda}{2} \|w\|^2$, weight decay term prevent network from over fitting.

3. Empirical Verification

Figure 1 shows that it can achieve 95% accuracy on the testing set.

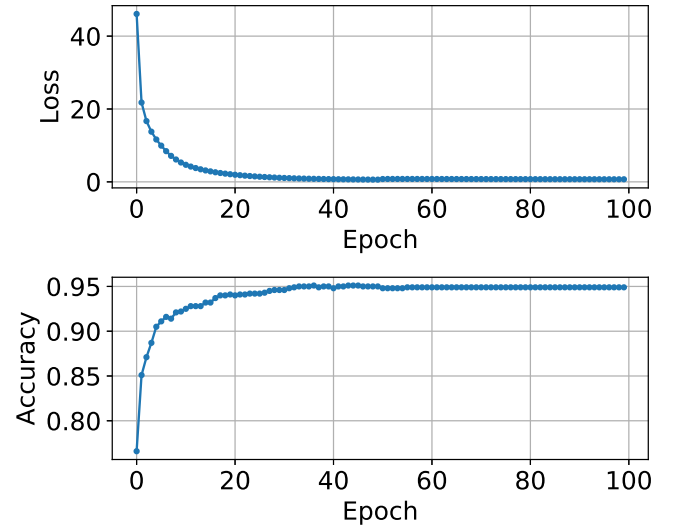


Figure 1. Loss and Accuracy

References

- [1] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, vol. 9, pp. 249–256, 2010.
- [2] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, *Proceedings of the IEEE*, 86(11):2278-2324, November 1998,