COEN 240 Machine Learning

Homework #3

Guideline: Please complete the following problems and generate a PDF file. Please submit the PDF file and a separate zip file that contains all source code to Camino. Please refer to HomeworkFormat.pdf for the format of the submitted PDF file.

Problem 1: The MNIST data set is consisted of gray-scale images of hand-written digits. Each image has 28×28 pixels. There are 10 classes, that is, digit 0, 1, 2, …, 9. There are 60,000 training images and 10,000 test images. The goal is to recognize the digit on the image. Use multi-class logistic regression for the hand-written digit recognition task. Give the recognition accuracy rate, and show the confusion matrix. The following code segment is for your reference:

```
import tensorflow as tf
import numpy as np
mnist = tf.keras.datasets.mnist
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
(x_traino, y_train),(x_testo, y_test) = mnist.load_data()
x_train = np.reshape(x_traino,(60000,28*28))
x_test = np.reshape(x_testo,(10000,28*28))
x_train, x_test = x_train / 255.0, x_test / 255.0
logreg = LogisticRegression(solver='saga', multi_class='multinomial',max_iter = 100,verbose=2)
```

Problem 2: Build a two-layer neural network for the hand-written digits recognition task with the MNIST data set. The hidden layer has 512 nodes, and adopts the ReLU activation function; the output layer has 10 nodes, and adopts the softmax activation function. Use the cross-entropy error function, and run 5 epochs. Give the recognition accuracy rate and show the confusion matrix, both for the test set.

Problem 3

Consider a neural network that has K output nodes. The error function adopted at the output layer is the sum-squared-error cost function $E_n = \frac{1}{2}\sum_{k=1}^{K}(y_{nk} - t_{nk})^2$, where $y_{nk} = y_k(\mathbf{x}_n)$ is the k-th output of the n-th data sample, and t_{nk} is the k-th target value of the n-th data sample. The output nodes adopt the sigmoid activation function.

a. Derive the math expression of δ_k for the k-th output node.

b. A local structure of this neural network is shown in Figure 1. Assume that the right-most layer in the figure is the output layer, and you have obtained δ_k, k=1, 2, …, K in question a.

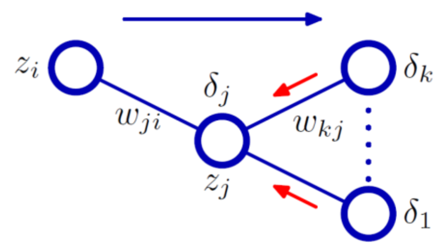

Figure 1.

Derive the math expression of δ_j for the j-th node in the middle layer in the figure, that is, one preceding layer of the output layer. Assume that the middle layer adopts the tanh function as the activation function h. Use the result you obtained in question a.