Spring 2023

CSE432/532 Machine Learning Basics

# Reading HW for Ch4.4

General Instructions:

1. Please carefully read the "*Chapter 4.4 Neural Network Demos V2.pdf*" and the shared code to answer the following question.
2. Please submit a ".pdf" file to the canvas. Typing is highly recommended.
3. You are encouraged to use any tools to finish this homework, including ChatGPT. You should not directly copy the answer from others.
4. Please submit the homework by next Thursday (March 30, 23:59PM).

# Q1 Before you start(5pts):

1. Make sure you are reading the slide "Chapter 4.4 Neural Network Demos V2.pdf" from Canvas. Please click the link in Page 5 to read the related code. You can also open a new session in ChatGPT for help.
2. Please read page 4 and briefly describe the general workflow for machine learning applications.

   The general workflow of machine learning applications typically includes these steps:
   1. Collect and clean/prepare data based on the requirements of the problem and the model.
   2. Choose an appropriate machine learning model for the task.
   3. Train the model with prepared data to learn the possible underlying relationships.
   4. Test/Evaluate the results from training the selected model.
   5. Tune the model based on the testing results to improve the performance.
   6. Repeat step 3 to 5 to optimize the performance.

# Q2 Problem & Data(15pts):

1. Please read pages 5-9 and the related coding part (the first step).
2. Can you briefly summarize what is the problem you want to solve and why we should use machine learning algorithm instead of traditional algorithm to solve this type of problem?

   - We are given a set of 2-D data points, pre-classified to 3 different labels. We have to come up with an algorithm to classify any given 2-D data point to one of the 3 labels.
   - We should use machine learning algorithm instead of traditional algorithms because such algorithm either does not exist, or would be so complicated the effort required to implement it would not justify the supposedly slight improvement in performance compared to the performance of machine learning algorithms.

3. What is the dataset used in this demo?

In this demo, we used a random generated dataset of 300 datapoints in 3 classes, 100 datapoints each. Let 0, 1, 2 be the labels. For each label j, its 100 datapoints are generated by the following code:

```
ix=range(N*j,N*(j+1))
r=np.linspace(0.0,1,N)
t=np.linspace(j*4,(j+1)*4,N)+np.random.randn(N)*0.2
X[ix]=np.c_[r*np.sin(t),r*np.cos(t)]
Y[ix]=j
```

So each point in the dataset is generated by taking a value r between 0 and 1, and transforming it into x and y coordinates using sine and cosine functions of t. t is an array of N values between j*4 and (j+1)*4, with each value perturbed by a small amount of Gaussian noise using np.random.randn(N) * 0.2. The addition of Gaussian noise ensures that the points are not perfectly aligned on the circle and introduces some variation in the dataset.

4. Please leverage the ChatGPT to understand dataset/data/label. Then, please explain what is the label in this dataset. How is the label stored in the code?

The labels in this dataset are represented by the Y variable, which is a 1D numpy array of length 300 containing integers between 0 and 2.
The pairs 0-99, 100-199, 200-299 in array X is the coordinate of datapoints of label 0, 1, 2 respectively. In this Y, element 0-99 are 0s, 100-199 are 1s, and 200-299 are 2s.

# Q3 About the model (30pts):

1. Please refresh your mind on what we have learnt so far and find the information from ChatGPT to explain what is the model in machine learning and what is the model in the first demo (Page 11-13).

- In machine learning, a model is a mathematical representation of a system (somewhat an equation) that is used to make predictions or decisions based on input data.
- In the first demo, the model is softmax(X * W + b) where X is the datapoints, W is a parameter (a small number), and b is the bias vector.

2. How do we define the model in the code?
   - Firstly, we initialize the parameters that will be used in the model

   ```
   W=0.01*np.random.randn(D,K)
   b=np.zeros((1,K))
   ```

   - In this code, instead of defining the model as a function that takes in parameters and call it whenever needed, we explicitly write code to execute step by step of the model. During the calculation of forward propagation, the code is written like so:

   ```
   #forward propagation
   #y=(x*w+b)
   h1=np.dot(X,W)
   h2=b+h1
   #softmax()
   exp_scores=np.exp(h2)
   s=exp_scores/np.sum(exp_scores,axis=1,keepdims=True)
   ```

3. Please read through page 12-17 and briefly explain what should be included in the computational graph based on Page 17.

<span style="color:red">Based on page 17, the computational graph should include the initial variables and parameters, as well as all the operations that are used to compute the scores, probabilities, and the loss function. The computational graph should also include the updates to the weights and biases using the gradients and the learning rate.</span>

4. Please read the pages 20 -38 and explain why do we have to carry out the forward propagation first or can we carry out the backward propagation without forward propagation?

<span style="color:red">We can not carry out backward propagation without forward propagation. Because calculating backward propagation require the values of local gradients, which requires the values calculated during forward propagation.</span>

5. Read the code and play with it. Then please finish the following table by filling with the size of the corresponding variable. For example, for W, its size (or shape) is (2,3).

| Variable | size | Variable | size | Variable | size |
|----------|------|----------|------|----------|------|
| W | (2,3) | H2 | <span style="color:red">(300, 3)</span> | dh2 | <span style="color:red">(300, 3)</span> |
| b | <span style="color:red">(1,3)</span> | S | <span style="color:red">(300, 3)</span> | dh1 | <span style="color:red">(300, 3)</span> |
| X | <span style="color:red">(300, 2)</span> | l1 | <span style="color:red">() or a number</span> | dW | <span style="color:red">(2, 3)</span> |

| Y | (300,) or (300, 1) | I2 | () or a number | db | (1, 3) |
|----|----|----|----|----|----|
| H1 | (300, 3) | I3 | () or a number | loss | () or a number |