Weicong Feng

Professor Konstantinos Krampis

BIOL79303

09/20/2021

Assignment3

Assignment 3

Machine Learning and Artificial Intelligence

Due date: Wednesday, 09/22/2021

For the questions below either write your answer in a few sentences or do the requested math. 10 points each question.

Question 1.

Which of the following are NOT parameters (based on the definition of a "parameter" from the text): weights, inputs, bias, constant e, learning rate?

Ans: inputs and constant are not parameters and learning rate should be called superparameter.

Question 2.

What is the difference between the logit and the logistic regression? Can the logistic function be a logit function?

Ans: In the textbook, “Introduction to Deep Learning”, the logit means the first part computation of logistic regression, , while the logistic regression is a model, including the logit and logistic function .

Logistic function is a non-linear function, while logit function is a linear function. So, the logistic function can’t be a logit function.

However, in other scenario, the logit function refer to the invert function of the sigmoid function, logit(x) = log(x / (1 – x)).

Question 3.

For the input and weight vectors (last item in the vector is the label), calculate the sum of squared errors, within a single "round" of a perceptron neural network:

xa = (0.5,1.2,1), xb=(1.2,1,1), xC=(0.3,0.4,0), w=(3,4), bias b=0.8

Note: use the standard logistic function as the logit function.

Ans: , ,

So,

Question 4.

Define what an epoch is. What happens during a single epoch, which parts of the neural network change? When would you think, we should stop running addition epochs?

Ans: In machine learning, an epoch refers to the number of passes of the entire training dataset for the machine learning algorithm. During a single epoch, in deep learning, the data instances are feed into the model, the forward-propagation procedure will calculate the loss function, and the backpropagation is used to update the weights, all the bias, weights in each layer and each neuro. When to stop the training dependents on the strategy. Sometimes, we can indicate a specific fix number of epochs; a common strategy called early stop indicates a specific metric, such that the training will stop when the model reaches the metric.

Question 5.

Write in matrix format, two epochs of the training of the feed forward neural network shown below, with the input data:

Xa = (0.6,0.4,2,0) Xb = (1,0.8,1,0) Xc = (1.2,0.7,1.5,1) , weights w= your choice and bias = 0.8

Your choice of weight adjustment after the first epoch, should be based on the predicted label / actual label difference (your choice should be wise to increase or decrease the predicted label accordingly).

Also calculate the sum of squares error. Do you observe it increasing or decreasing based on you weights choice? Also use the logistic function for the logit.

Diagram

Description automatically generated

The weights in layer 1 are initiated as

So, during the first epoch,

Suppose the weights in output layer are initiated as:

,

So,

Then,

Backpropagation will be used to calculate respect to each weights, and will be used to update weights for next epoch. Suppose that,

Suppose the weights are updated as below:

,

It is larger than the first epoch.

Question 6.

How does a computer "see" an image? (Explain the data which form an image).

What are the different approaches to treat the different color layers in an image? In your opinion which one is the best / you would choose and why?

Ans: Usually, there are two major approaches to represent a RGB image in machine learning. One is row x column x (R, G, B), which means it do this like that of gray image, and the only difference is it uses 3-dimension vector to stand for red, green and blue color in each pixel instead of a 0-255 value. Another way is color channel x row x column, that the color channel is 3, which represents red, green and blue respectively.

I prefer the latter because it is easy to feed into the neuro network.

Question 7.

Is K-means supervised method, and what are the centroids? Do the centroids move during the cycles of the K-means?

Ans: the centroids are the central points in each cluster. The centroids will be recalculated and reassigned in each cycle.

Question 8.

In the Principal Components Analysis (PCA) do the data points change? Remember a data point is just coordinates (2 coordinates in 2D space, 3 in 3D and so on). What does it mean to have "correlated variables”? Give an example different from the book. Does the variance increase or decrease during rotation and how?

Ans: In the PCA, the data is changed in different coordinate system. It is called transformation. Correlated variables refer to the variables are linear dependent in each other. It means that one or more variables can be presented as linear combination of other variables.

Suppose we have following data set:

Chart, bubble chart

Description automatically generated

Through PCA transformation, it actually rotates the coordinate system, the data can be shown below:

Chart, scatter chart

Description automatically generated

The data set can be presented by only one dimension without losing much information.

Question 9.

What is the difference between the bag of words representation and the one-hot encoding?

What is an example of non-linearity function (the logit function), that we mentioned in the class in addition to the logistic function? Can you have different non-linearity functions within the same layer? If yes, are there any potential problems with this?

Ans: bag of words is an algorithm expands words in corpus to features which count the number of times encounter in the corpus, so it could be greater than 1. However, one-hot is a technique which encodes a categorical variable to multiple features, that all is 0 except of the corresponding feature being 1. Hence, inside one-hot encoding, the value is either 0 or 1.

Besides Sigmoid function, there are some alternative non-linear functions can be used as activation function, for instance, Tanh and ReLU.

I think it is possible to add more than one non-linear function into same layer because we can have multiple branches in one layer. If multiple non-linear functions are stacked into one layer without branch, it depends on the kind of non-linear function. For example, feeding the output of a sigmoid function into a tanh will lead to the magnitude become too tiny. However, stacking ReLU on a tanh works, but why don’t use sigmoid instead? In addition, I found a paper (“[Learning Neural Networks with Two Nonlinear Layers in Polynomial Time](https://par.nsf.gov/servlets/purl/10112755)”, Surbhi Goel and Adam R. Klivans) whose author created a model has two non-linear functions in a layer.

Question 10.

What operations take place during the forward pass of the feed- forward neural network, and what do we mean by "function composition”?

Ans: during the forward pass of the feed-forward neural network, the input was feed into the network, the logistic regression in each neuron is calculated, in which x multiply the weights, sum up together, then plus b. The results are passed to activation, and output to next layer as input. After going through all layer, the predicted result is going to be compared with the label by using loss function.

The logistic regression is function composition, .Between layers,

.