Weicong Feng

Professor Konstantinos Krampis

BIOL79303

09/05/2021

Assignment2

Assignment 2

Due date: Wednesday, 09/15/2021

QUESTION 1: Which data structure do we call as "graph of a function"? Give 2-3 examples with multidimensional input data (3,4 and 5 dimensions), along with the labels of 1 and 0.

The graph of a function is a graphical expression of the functions inputs and f(x).

For example:

|  |  |  |  |
| --- | --- | --- | --- |
| Long | Weight | Length of Tail | Dog/cat |
| 28 | 60 | 12 | 1 |
| 10 | 12 | 4.2 | 0 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Attendance | Assignments | Project | Quiz | Pass Exam |
| 10 | 88 | B | 80 | 1 |
| 6 | 65 | B | 40 | 0 |

QUESTION 2: What do you think a discontinuous function would be, partial or total? (based on the definitions we see in the written material). Does it affect the continuity of a function if it is surjective or injective?

Ans: A discontinuous function would be a partial function because one or more elements in the domain of a discontinuous function isn’t defined. In term of surjection, any element in function codomain has at least one element in domain which are mapped to it, so it must be continuous; however, for the surjective function, any element in function codomain has at most one element in domain which is mapped to it, so it may be continuous or discontinuous; similarly, bijective function must be continuous.

QUESTION 3: Find the derivatives of the following functions, using a combination of the rules, analytical approaches seen in the text and algebraic manipulation. Please include in your answer the complete steps of how you reached the derivative equation.

Ans: 1)

2)

Set , then

3)

.

QUESTION 4: Write the 3-dimensional vector (2,5,4) as a linear combination of the standard basis.

Additionally, answer whether we can create a basis B with vectors (2,1) and (1,4)? TIP: you need to think of the definition of the basis, and use the formula for linear combination of vectors in the corresponding section of the book (the formula is for 3 dimensions, but you can do the same for 2)

Ans: (2, 5, 4) = s1e1 + s2e2 + s3e3 = s1(1, 0, 0) + s2(0, 1, 0) + s3(0, 0, 1)

= 2(1, 0, 0) + 5(0, 1, 0) + 4(0, 0, 1)

If (2, 5, 4) can be represented by a basis B with vectors (2, 1) and (1, 4), then there must be:

1. (2, 1) and (1, 4) are independent.

So, 2a + b =0 and a + 4b = 0

Solve this equation system, it has only the trivial solution. Therefore, (2, 1) and (1, 4) are independent.

1. 🡺

So, . The 3rd row tells that there is no solution in this equation system.

Hence, vectors (2, 1) and (1, 4) are not the basis of vector (2, 5, 4).

QUESTION 5: Find the gradient (write the vector of all the partial derivatives) of the following functions:



TIP: Get each unknown in the function and setting the other as constant get its partial derivative.

Ans: 1)

2)

QUESTION 6: Explain what overfitting is, and the problem if we build a Machine Learning model that is overfitted to the training data. For the following data vectors, build a table that is one- hot encoded for the categorical data (the last element of the vectors is the label of the data point).

(15,20, Long,Crane)

(11,20, Short, Sparrow)

(5,30, Medium, Pigeon)

(16,22, Medium, Crane)

(9,15, Short, Sparrow)

Ans: When the models are trained too much closing to the training data, and testing errors is relative bigger than the training errors, the models are overfitting. In that case, we should try to generalize the models, for example, trying simpler models, normalization or more data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Weight | Height | Long | Medium | Short | Target |
| 15 | 20 | 1 | 0 | 0 | Crane |
| 11 | 20 | 0 | 0 | 1 | Sparrow |
| 5 | 30 | 0 | 1 | 0 | Pigeon |
| 16 | 22 | 0 | 1 | 0 | Crane |
| 9 | 15 | 0 | 0 | 1 | Sparrow |

QUESTION 7: Below is the confusion matrix (following the same structure as in the text) for a hypothetical model we have just trained. Calculate precision and recall. \*\* Important note \*\* There is an error in the confusion matrix as it is shown in the book text (hence preventing you from calculating the precision because of a missing term).

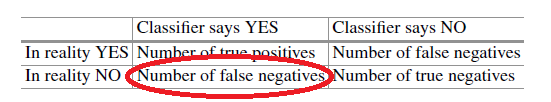
|  |  |  |
| --- | --- | --- |
|  | Predicted Positive | Predicted Negative |
| Actual Positive | 10 | 3 |
| Actual Negative | 2 | 8 |

Find where the error is by looking carefully at the definitions in each of the 4 cells of the matrix (there are only four cells with the metrics, the other with YES/NO are headers providing the definitions).

Precision =

Recall = = 0.769

In the textbook, the confusion matrix is shown as below:



The cell being in red oval is wrong, the item that actually is negative but is classified as positive is called false positive.

QUESTION 8 (this is from chapter 3, section 3.4 in the Skansi book for which I added an additional recorded lecture): It is mentioned that logistic regression is a “one neuron network", but does it actually have any restrictions on how many dimensions the input data can have, whether you choose yes or no, please justify.

Follow the two cycles of calculation in section 3.4 (where after the first cycle it adjusts the weights and repeats the calculation). What do you observe in the output / predicted y values compared to the actual data, do they converge?

You task is now to perform one more round of calculation, by adjusting the weights, and redoing the sums and final calculation with the logit function (for the calculation of exponent of the Euler number in the logit function, you can use any scientific calculator or there are plenty virtual calculators on the internet).

Note the book mentions “magic” in adjusting the weights (which is the backpropagation algorithm which we will cover in a couple of lectures). What you need to do is just select some weights based on you intuition which you think will reduce the error or the output y compared to the real y).

\*\* Most important \*\* I would like you to do the calculation using the matrix/vector notation to represent the inputs and the weights, as shown in later in section 3.4 of the Skansi book (pages 66-67) and show the calculations in matrix format (the logit function will be finally applied to all elements of the resulting vector with the sums).

Ans: No restriction on how many dimensions the input data can have in a “one neuron network.” The computation inside deep learning neuron is a logistic regression, which can contain no mater how many weighted input elements you want, as shown below:

and

Let’s say, we have a simple neural network with 4 inputs:

b

Z

The training data set we have:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x1 | x2 | x3 | x4 | Label |
| 0.7 | 0.2 | 0.8 | 0.15 | 0 |
| 0.9 | 0.86 | 0.2 | 0.78 | 0 |
| 0.2 | 0.05 | 0.25 | 0.02 | 0 |
| … | … | … | … | … |

In the first epoch, suppose patch size is 3, a random initial weights matrix is created:

So,

The cost function

= 0.38

Next, it is gradient descent:

Similarly,

-0.26

-0.48

-0.21

If the is set as 1, the weights are updated as below:

After updating weights, the logistic regression equation becomes:

Z = 1.145 + 0.64x1 + 0.86x2 + 0.78x3 + 1.01x4

Hence, in the second epoch:

So,

The network coverge.

QUESTION 9: Let’s say we have a “MNIST2” dataset which has 56x56 size images. How long would be the vector, if we convert an image into a vector?

Look back through the definitions in the chapter, where we saw mentioned a “tensor”. If we have colored rectangular image with 1 Megapixel (1 million pixels), what would be the resulting dimensions of the tensor from this image?

Ans: When a 56x56 gray image is converted to a vector, it would be a 3136-dimensional vector. A 3-channels 1 Megapixel color image can be converted to a 1,000,000 x 3-dimensional tensor.

References:

Skansi, Sandro. *Introduction to Deep Learning*. Springer, 2018.

Pandey, Parul. *Understanding the Mathematics behind Gradient Descent*. Towards Data Science,

March 18, 2019

<https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e>. Accessed September 11, 2021.