Weicong Feng

Professor Konstantinos Krampis

BIOL79303

11/12/2021

Assignment9

Assignment 9

Due date: Wednesday, 11/17/2021

**Machine Learning and Artificial Intelligence for Bioinformatics**

**Each of the questions below is worth 10 points.**

**\*\* This homework examines also material posted for the October 27th and November 3rd lecture \*\*.** Please read those past materials as well (if you have not done so already), both the current and those earlier materials are needed to answer the questions below.

Simply write out each answer in 2-3 short sentences. If the question requires you to write math formulas, you can use Latex or write them by hand, but make sure you have clear writing and the pictures you get of the writing are also clear and aligned to the rest of the document.

**Question 1.** Describe how do you know that a univariate function has minima, maxima, or none at all? Is it a requirement for the function to be continuous, in order to have any of these?

*Ans: The derivative of a univariate function with respect to its independent variable would be zero at maxima and minima. Set the first derivative f’(x) = 0, if its second derivative of f’’(x) is positive, that is a minima; while the f’’(x) is negative, that is a maxima; if f’’(x) turns out to be zero, then the point is called a point of inflection, none minima or maxima. In order to there exist these stationary points, the function f(x) must be continued and defined in these points.*

**Question 2.** What is a Hessian matrix of a function? How is the Hessian related to the gradient of a function?

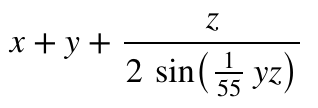
*Ans: The Hessian matrix is a symmetric matrix containing all the second derivatives of the multivariate function (*[*https://www.sciencedirect.com/topics/mathematics/hessian-matrix*](https://www.sciencedirect.com/topics/mathematics/hessian-matrix)*). For f(x, y, z), it should be*

*Text, letter

Description automatically generated*

*The Hessian matrix of a function f is the Jacobian matrix of the gradient of the function f.*

**Question 3.** Please answer how many entries the Hessian Matrix of the following function has (just report the number of entries, no need to get the derivatives, but feel free to present them if you want):



*Ans: There are 9 entries in the Hessian Matrix of above function.*

**Question 4.** Read page 27 of Chapter 1 in file “Single and Multivariable Calculus - Hands-On Math for Deep Learning”, what do the terms 𝑓*xx*  , 𝑓*xy* , 𝑓*yx*  , 𝑓*yy*  stand for ?

**Question 5.** Similarly on page 27 of Chapter 1 in file “Single and Multivariable Calculus - Hands-On Math for Deep Learning”, describe the process of finding the stationary points (minima, maxima) of a multivariate function.

**Question 6.** Read the “second derivative test” section from <https://en.wikipedia.org/wiki/Hessian_matrix> (feel free to read any additional material from that page if you want to). List the different cases of Hessian matrices and how these indicate whether the function has minimum, maximum or saddle points?

*Ans: For the Hessian matrix:*

*Positive-definite at x: local minimum at x*

*Negative-definite at x: local maximum at x*

*Both: saddle point at x*

*Otherwise: inconclusive*

**Question 7.** What is the difference between convex and non-convex function, and why would you prefer one versus other as loss function used in an artificial neural network?

*Ans: f is called convex if and only if any of the following equivalent conditions hold:*

*For all 0<=t<=1 and all :*

*f(tx1 + (1-t)x2) <= tf(x1) + (1-t)f(x2)*

*Artificial neural network prefer convex function because a strictly convex function on an open set has no more than one minimum.*

**Question 8.** Define and give an example of “drop-out” in artificial neural networks. What is drop-out used for and how is it beneficial in training the neural network?

**Question 9.** List the different methods for regularization. Describe how the comparison of the training data vs testing data error, can be used to prevent overfitting of the data? Read the “drop-out” section, and shortly describe in your opinion, how it can be used in the case of training a hand-writing digit recognition dataset (like MNIST), which has digits written only 2-3 persons (so they are very similar) \*\* see footnote.

**Question 10.** Define what the Delta rule is, and a short description on how we reach the Delta rule (what is the starting formula, and what steps we take on it - do not list the formulas, explain / describe the steps). How does the Delta rule relate to backpropagation? What is the problems backpropagation solving, in comparison to doing manual weight updates?

\*\*Remember, different neurons in a fully connected network learn to recognize different micro-patterns in the digit images. So if the images are very similar written by a few people, what would happen, and how can this be prevented with drop-out ?