Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNIA Winter Term 2019/2020

Exercise Sheet 5

(Solutions)

Deadline: 15.12.2020, 23:59

Instructions

Submit the jupyter notebook with the solution for exercise 5.2 b) in an archive along with the latex file.

Exercises

Exercise 5.1 - Computing Jacobian and Hessian (1 + 1 = 2 points)

Let $f(x, y) = 3x^2y + 4x^3y^4 - 7x^9y^4$. Compute Jacobian and Hessian matrices of f.

Solution 5.1

jocobian: $\begin{bmatrix} 6xy + 12x^2y^4 - 63x^8y^4 & 3x^2 + 16x^3y^3 - 28x^9y^3 \end{bmatrix}$

Hessian: $\begin{bmatrix} 6y + 24xy^4 - 504x^7y^4 & 6x + 48x^2y^3 - 252x^8y^3 \\ 6x + 48x^2y^3 - 252x^8y^3 & 48x^3y^2 - 252x^9y^2 \end{bmatrix}$

Exercise 5.2 - Taylor Series and Newton's Method (1 + 2 + 1 + 1 = 5 points)

a) Derive the first 5 terms of the Taylor series about $x_0 = 0$ for $f(x) = cos(x)$, and write the series in sigma notation (e.g. as an infinite sum).

b) In python, apply Newton's method to find the nearest critical point of

$$f(x, y) = x^2 - y^2 + 4 - 3xy$$
from the initial point $x_0 = -0.3, y_0 = 0.3$.

After each iteration, check the value of the first derivative, i.e. Jacobian: if Jacobian is 0, then we reached the critical point.
Plot the original function for x and y in range from -0.5 to 0.5 with step size of 0.01, along with the initial point and the points computed after each iteration. Use method .surface_plot() with parameter $alpha = 0.3$ for plotting the function and .scatter() for plotting the points.
What kind of problem of function minimization task is illustrated with this example?

c) How is Newton's Method related to gradient descent?

d) In which case is it impossible to apply Newton's method? Hint: look at the multidimensional generalization of the formula.

Solution 5.2

a)
$1,0,-\frac{x^2}{2},0,+\frac{x^4}{4!}$
$\cos x = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n}}{(2n)!}$
b) see 5.2b.ipynb
c)
d) when Hessian metrices is not invertible

Exercise 5.3 - Activation Functions (1.5 + 1 + 0.5 = 3 points)

a) Three of the most commonly-used activation functions are the sigmoid function, hyperbolic tangent, and ReLU. The equations for these are given below. Compute the first derivative of each function. Note that your final derivative for tanh should not be written in terms of other hyperbolic functions, though you may use these in your calculation. Hint: ReLU is not differentiable at x = 0. For the purposes of your derivative, you may define its derivative piecewise, ignoring this point.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad \tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$$

$$ReLU(x) = \begin{cases} 0 & x < 0 \\ x & x > 0 \end{cases}$$

b) Using an online resource like Wolfram Alpha or Desmos, graph each function along with its derivative. Discuss the differences you observe. What are the advantages and disadvantages of each? In particular, think about how the range of the function and the amplitude of the derivative would affect a network.

c) Which activation function would be most appropriate for a classification problem when there are only two classes? Would adding more classes change your choice? Why or why not?

Solution 5.3

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- You have to submit the solutions of this assignment sheet as a team of 2-3 students.
- Hand in a **single** PDF file with your solutions.

- Make sure to write the student teams ID and the name of each member of your team on your submission.
- Your assignment solution must be uploaded by only **one** of your team members to the course website.
- If you have any trouble with the submission, contact your tutor **before** the deadline.