Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNIA Winter Term 2018/2019

Exercise Sheet 8

Computational Graphs and Regularization

Deadline: 07.01.2019, 23:59

Exercise 8.1 - Computational Graph (1 + 2 = 3 points)

In the Exercise 6.3 part (b), you have applied a forward and backward pass to compute the derivative of a term represented by a computational graph.
In this exercise we want to do the same with a logistic regression classifier $f(x) = \sigma(x_1 w_1 + x_2 w_2 + b)$ with the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$.

a) Setup the computation graph of the classifier. Split the whole model term into simple operations (i.e. Add, subtract, multiply, divide and exponential) that form the nodes of the graph. For each operation, state its derivative.

b) Perform a forward and backward pass using the following values:

$x_1 = 1, x_2 = 0.5, w_1 = 0.25, w_2 = 0.3, b = 1$

Keeping all other weights fix, how must w_1 be adapted in order to flip the classification result of the given input? Justify your answer.

Exercise 8.2 - ReLU and Tanh Functions (1 + 1 = 2 points)

In Exercise 6.2. we calculated and sketched the derivative for sigmoid function, we will do the same for other activation functions to understand the difference between them and the benefit of each in Neural Networks:

a) Calculate the derivative of the tanh and ReLU function by hand.

b) Sketch the gradient of the tanh, ReLU functions and also the sigmoid you sketched in the Exercise 6.2, all in the same graph (please indicate ticks on the axes).

Now explain what do you observe from this graph (i.e what are the differences between activation functions' gradient, which one would be more suitable for back-propagation and which one would create more problems and why) ?

Exercise 8.3 - Lagrange Multiplier (2.5 points)

Lagrange Multiplier is a widely used method for optimizing functions under constraints. You can read more about it on https://en.wikipedia.org/wiki/Lagrange_multiplier.

Find critical points of the function $f(x, y) = x^3 + xy^2$ under the constraint $2x + y^2 = 2$ using Lagrange Multiplier. (No need to specify whether they are minimum or maximum)

Exercise 8.4 - Ridge and Lasso $(0.5 + 0.5 + 1.5 = 2.5$ points)

The p-Norm is defined as

$$\|\mathbf{w}\|_p = \left(\sum_{i=1}^{n} |w_i|^p\right)^{\frac{1}{p}} \qquad \textit{where } p \geq 1 \textit{ real number.}$$

a) Sketch a contour plot of L_p norm for $p \in \{1, 2\}$.

b) From the previous sketch, What are the main advantages and drawbacks of using lasso instead of ridge regression? (Hint: think about differentiability and the effect on weights)

c) In the lecture, Ridge Regression was expressed as an unconstrained optimization problem: $min_w \left(\frac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2\right)$.

   Show that this is equivalent to the constrained optimization problem: $min_w(\frac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2)$ constrained by $\|\mathbf{w}\|_2^2 \leq s$.

   What is the relation between λ and s ?

   Hint: Use Lagrange Multiplier (Solve Exercise 8.3 first)

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 a solution of this assignment sheet as a team of 2-3 members.
- Hand in a **single** PDF file with your solutions to the tasks.
- Therefore, make sure to write the name and matriculation ID of each of the members in your team.
- The 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.

Plagiarism of any form is not tolerated. If you refer something from the web, you must give proper credit by citing the source. Lack of this would be considered plagiarism. In such a case, the whole sheet would be awarded zero points and a warning is given. If this act is repeated again, then the whole team is excluded from the course.