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

Exercise Sheet 6

Deadline: 10.12.2018, 23:59

Exercise 6.1 - Maximum Likelihood Estimation and Cross-Entropy (0.5 + 0.5 + 1 + 2 = 4 points)

Given a set of m i.i.d. samples $\mathcal{X} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\}$ drawn from a data-generating distribution $p_{data}(\mathbf{x})$ and a parametric family of probability distributions over the same space $p_{model}(\mathbf{x}; \theta)$.

a) Write down the maximum likelihood estimator for θ.

b) Explain the difference between empirical distribution $\hat{p}_{data}(\mathbf{x})$ and the data-generating distribution $p_{data}(\mathbf{x})$.

c) Rewrite the expression derived in a) as an expectation using the empirical distribution $\hat{p}_{data}(\mathbf{x})$. Give an argument for why this is possible.

d) Show that minimizing the cross-entropy between $\hat{p}_{data}(\mathbf{x})$ and $p_{model}(\mathbf{x}; \theta)$ is exactly the same as computing the maximum likelihood estimator in a).

Hint: *Note that the definition of KL-divergence is*:

$$D_{KL}(P||Q) = \mathbb{E}_{x \sim P}\ [log\ \frac{P(x)}{Q(x)}]$$

Hint: *For this question, it might be useful to refer to the introduction of the optimization chapter in Ian Goodfellow's deep learning book.*

Exercise 6.2 - Sigmoid Function (0.5 + 1 + 1 + 1.5 = 4 points)

The commonly used activation function in hidden layers of a Neural Network is the Sigmoid function which is defined as:

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

a) Prove that the derivative of the sigmoid function is $\sigma(x) - \sigma^2(x)$.

b) Sketch the gradient of the sigmoid function (please indicate ticks on the axes) and also explain what are the inherent properties that you observe from the above computed gradient?

c) Prove that the sigmoid function is point symmetric.

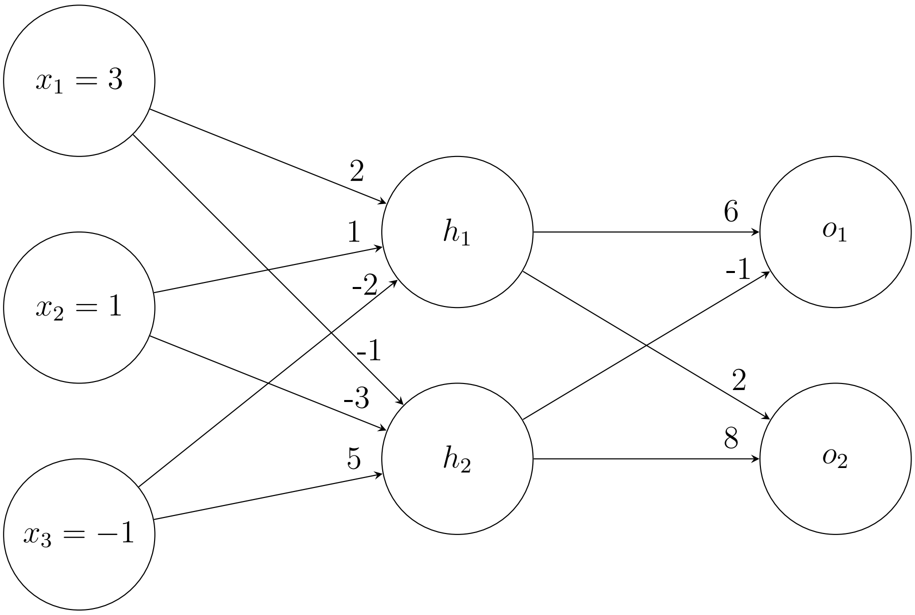
Hint: You can check the following wikipedia page: (https://en.wikipedia.org/wiki/Point_reflection) for point symmetric meaning.

d) We know from Newton's method the importance of Taylor series in optimization, additionally, Taylor expansion could be beneficial in providing a cheaper computation alternative for activation functions (for further reading: http://www.yildiz.edu.tr/~tulay/publications/Tainn2003-3.pdf).

So, **find** the first 3 terms in the Taylor series for the sigmoid function centered at 0.

Hint: You can use the derivative form proved in (a) when calculating higher derivatives.

Exercise 6.3 - Basics of Forward and Backward passes in computational graphs (1 + 1 = 2 points)

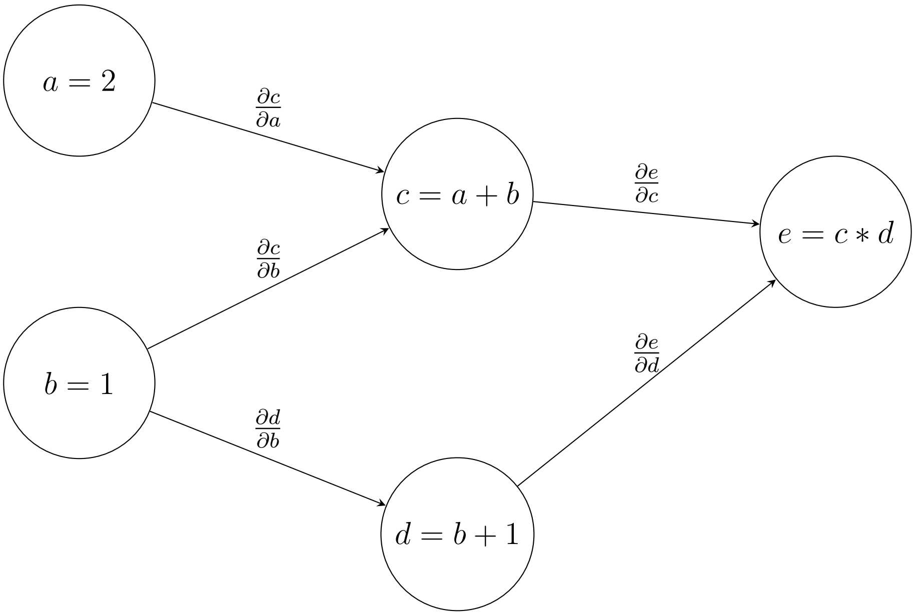


a) The simple one layer Neural Network (given above) takes an input of three features, and produces a vector output. Apply a forward pass with the given inputs and weights in the circles and above the arrows respectively, use ReLU function (ReLU function is defined as: $ReLU(x) = max(0, x)$) for the hidden nodes and softmax function for the output nodes.

If this is a binary classification problem, what would be the predicted class label for this given input. Hint: Think about what the output of the softmax implies.

b) For the computation graph given below, write down the expressions using chain rule and compute the final values for

- $\frac{\partial e}{\partial b}$
- $\frac{\partial e}{\partial a}$


$a = 2$
$b = 1$
$c = a + b$
$d = b + 1$
$e = c * d$
$\frac{\partial c}{\partial a}$
$\frac{\partial c}{\partial b}$
$\frac{\partial d}{\partial b}$
$\frac{\partial e}{\partial c}$
$\frac{\partial e}{\partial d}$

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.
- Therefore, make sure to write the student 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.