

**YΣ19 Artificial Intelligence II (Deep Learning for
Natural Language Processing)
Fall Semester 2020
Homework 2
20% of the course mark
Announced: October 27, 2020 Due: November 24,
2020 at 23:59**

1. Read Section 4.1.1 of the book *Pattern Recognition and Machine Learning* by Christopher M. Bishop (Springer, 2006, you can find it online easily). I have used material from this section in my lecture on perceptrons.

Prove equations (4.5) and (4.6) of that section, which are given by the author with no explanation. You can use any basic result from linear algebra.

(2/10 marks)

2. Let $\mathbf{x} \in \mathbb{R}^{1 \times n}$ be a row vector, $\mathbf{W} \in \mathbb{R}^{n \times m}$ be a matrix and $\mathbf{z} = \mathbf{x}\mathbf{W}$. Compute $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$.

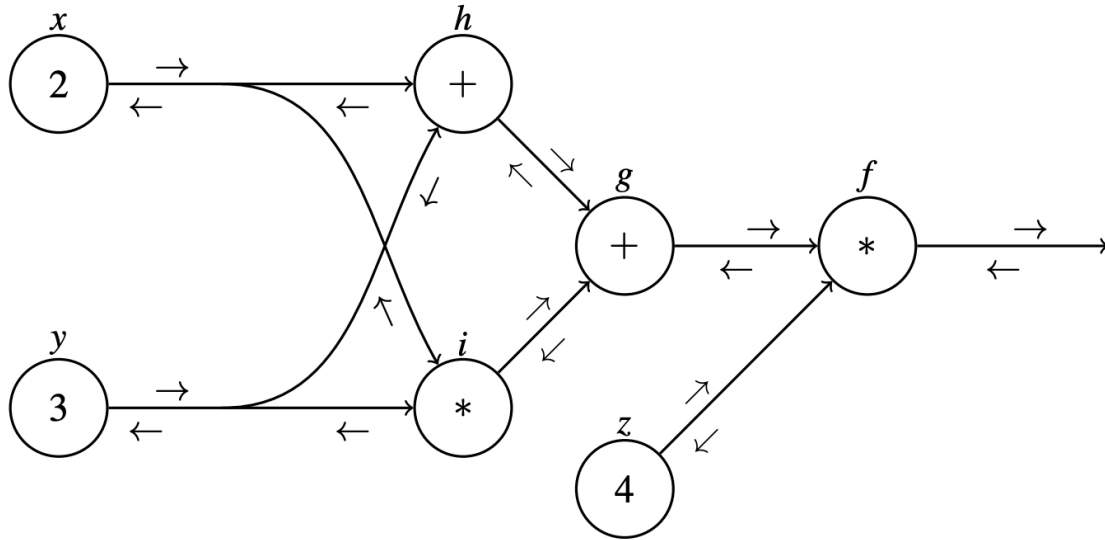
(1/10 marks)

3. Let $\hat{\mathbf{y}} = \sigma(\mathbf{x}^\top \mathbf{w})$ where $\mathbf{x}, \mathbf{w} \in \mathbb{R}^{n \times 1}$ are column vectors and σ is the logistic function (defined in the logistic regression lecture). Compute $\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{w}}$.

(1/10 marks)

4. Apply forward propagation and backward propagation to the following computational graph to obtain the values and the gradients missing from its edges.

(1/10 marks)



5. Develop a sentiment classifier using feed-forward neural networks for the Twitter sentiment analysis dataset of the previous assignment [available here](#). We expect you to experiment and develop more than one models (at least two) and compare them over F1 score, Recall and Precision. For the development of the models, experiment with:

- the number of hidden layers, and the number of their units
- the activation functions (only the ones presented in the lectures)
- the loss function
- the optimizer, etc

We encourage you to use fine-tuning techniques. Furthermore, in at least one of your models you must use pre-trained word embedding vectors using [GloVe](#).

In the end, choose your best model and describe the reasons why you concluded to that architecture, based on the theory you were taught. For the best model plot the loss vs epochs and the [ROC curve](#) and explain what you see. Also, compare it with the solution of your logistic regression from Homework 1. Your solution should be implemented in PyTorch and we expect your report to be well documented.

(5/10 marks)