Neural Networks (LTAT.02.001)

University of Tartu
Institute of Computer Science

Homework 3: Lecture Materials

- These theoretical questions are about the material covered in the lecture about "Back-propagation". You can get 5 pts for correctly answering these questions. These points are added to the 25 pts you can get from the coding task.
- All questions must be answered in your own words, do not copy-paste text or images from the internet.
- Submit the answers to these questions through courses.cs.ut.ee. Under the practice sessions you need to select "Homework 3:theory" and upload the file. The answers must be in a .pdf file. Points can be deducted for terrible formatting or incomprehensible English.

Back-propagation

Exercise 1: Derivatives of scalars, vectors and matrices (1pt)
In this task we ask you what are the dimensions of the gradient vector/matrix when taking the derivative of the object noted in the topmost row with respect to the object given in the first column. You can find help from http://cs231n.stanford.edu/handouts/derivatives.pdf

Please fill in the table below:

$\frac{\partial}{\partial}$	scalar	vector $\in \mathbb{R}^n$
scalar	$\cdots$	$\cdots$
vector $\in \mathbb{R}^k$	$\cdots$	$\cdots$

Exercise 2: General rule of backprop (1pt)
Imagine a feed-forward network with N hidden layers, without using any non-linearity in the nodes (for simplicity of this exercise). We note as W^N the weight matrix that is between layers N and N+1. We note as h_N the hidden node activations in layer N.
Considering that you already know the derivatives with respect to the hidden node activations in layer N+1, $\frac{\partial L}{\partial h_{N+1}}$, give the expressions for the following derivatives: $\frac{\partial L}{\partial W_N}$; $\frac{\partial L}{\partial h_N}$, $\frac{\partial L}{\partial W_{N-1}}$ and $\frac{\partial L}{\partial h_{N-1}}$.

Exercise 3: Complexity task 1 (1pt)
How many multiplications and how many additions/subtractions does it take to calculate the product of a vector $v \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{n \times m}$ (i.e. complexity of doing $v^T \cdot M$)?

Exercise 4: Complexity task 2 (2pt)
Imagine that you have a neural network with one hidden layer, using RELU activation function at the hidden nodes. The inputs to the network are RGB images of size 32x32x3. The output is 10 probability values (obtained by passing the results through Softmax function).
How many scalar operations (additions, subtractions, multiplications, divisions, exponentiations, logarithms and comparisons) does it take to

- perform a forward pass on one data point (to calculate the cross-entropy loss) in a network whose hidden layer contains 100 nodes.