Neural Networks (LTAT.02.001)
University of Tartu
Institute of Computer Science
Homework 4: Lecture Materials

- These theoretical questions are about the material covered in the lecture about "Optimization and regularization". 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 4: theory" and upload the file. The answers must be in a .pdf file. Points can be deducted for terrible formatting or incomprehensible English.

Optimization

Exercise 1: Optimization methods (3pt)
In the coding task you will be asked to implement multiple improvements to the naive approach to the stochastic gradient descent (SGD).
You can find the descriptions of the optimization methods at http://cs231n.github.io/neural-networks-3 .
Please find the following update rules from the document and explain in 1-2 sentences :

- SGD with momentum. Explain what is v. Why might it be useful to use v instead of dx when updating the values of x.

- RMSprop optimizer. Explain what the *cache* variable represents.

- Adam optimizer. Explain what is the difference with RMSProp.

Dropout

Dropout is a very important method for regularization, but we have no time to code it in the current practical session. Luckily, if interested, you can implement dropout as a bonus exercise (5pts).

Exercise 2: Dropout basics (2pt)

We are training a model with one hidden layer containing 10 hidden nodes. We use dropout on the nodes of that hidden layer, such that at each training step we drop out exactly 2 random nodes of that layer.

- over many (millions) training steps how many different models do we use (how many different ways of dropping out exactly two nodes there is)? Notice that the order of dropping nodes does not matter!

- When using this model at testing phase, with what constant do we need to multiply/divide the weights going out from this layer?