Lehrstuhl für Betriebswirtschaftslehre
insbesondere Wirtschaftsinformatik im Dienstleistungsbereich
Univ.-Prof. Dr.-Ing. Wolfgang Maaß

UNIVERSITÄT
DES
SAARLANDES

Homework 5
Introduction into Data Science (Summer term 2016)

Use "Homework_5.zip" to complete exercise below. Submit homework as a zip archive "Submission_5.zip". In your solution, exercise_0.txt, exercise_1.txt and exercise_2.txt should be reports to exercises.

Exercise 0 **(Installation of Keras, Theano in VM**) *20 points*

1.1. *In virtual machine of your choice (e.g. Virtualbox), install 64 bit Ubuntu OS 14.04. Please do not install the most recent Ubuntu. It is suggested that you allow your virtual machine to use all the cores of your CPU, to speed up computations in the exercises.*

1.2. *For the following steps, please save all the command line commands that you used to install the deep learning libraries and description of actions that you performed, such that you can reproduce the installation. Include this description as exercise_0.txt in submission.*

1.3. *After the installation of virtual machine, install Theano python library. Use the command line instructions provided on the official web - site to install it. Make sure that you follow the instructions for the right version of Ubuntu.*

1.4. *Install Keras python library for deep learning. Follow the instructions on the official website of the library.*

Exercise 1 **(Sentiment analysis**) 25 *points*

1.1. *The task in this exercise is to create a predictive model that can predict given the raw text contents of review on the IMDB website whether the review sentiment is positive or negative. Further description of the task as such is given in the lecture slides. Open the file imdb.py. This file comes from the examples of usage of Keras library. Get familiar with the contents of the file, in particular where the architecture of the deep neural network is defined and parameters of the optimization procedure.*

1.2. *The learning rate parameter determines how much a training procedure can change the weights of the neural network in a single training step. Try running the file with learning rate set to 0.001, 10.0, 0.000001. Report your observations of training and validation loss behaviors during the training of the neural network. Note that starting the computations with Theano + Keras might take some time, as Theano takes some time to compile the architecture of the deep neural network.*

1.3. *The number of neurons in the recurrent block of neural network determines how much modeling power a network can have. Try setting the number of neurons to 128 and running the training procedure with modified architecture. Report on how this affects best*

validation error. What is your expectation towards change in validation accuracy with further increase of number of neurons?

1.4. *Calculate and report the number of parameters in the training procedure and the model in the python script that you can adjust.*

Exercise 2 **(Computer vision)** 25 *points*

1.1. *The task in this exercise is to create a predictive model that can estimate given as input 32 x 32 color image which one out of 10 classes is depicted on the image. Open the file cifar10.py. This file comes from the examples of usage of Keras library. Get familiar with the contents of the file, in particular where the architecture of the deep neural network is defined.*

1.2. *Report on how labels for the images are represented for training of the deep neural network in cifar10.py.*

1.3. *Run training of the deep neural network with cifar10.py. Report the best validation accuracy you get. Is it better than the one with random guessing?*

1.4. *Set number of neurons in first fully connected layer of the network to 4. How does the validation error changes? Why do you think this happens? Write down answers to these questions into the report.*

deadline: Wednesday, 06/01/2016 24 pm