**Assignment 2 – Practical Deep Learning Workshop**

\* this assignment can be done in teams of up to 3 students

**Time series** – download the dataset and data dictionary from [here](https://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring).

1. Present an exploratory data analysis of the dataset.
   1. You can refer to the previous assignment, but also think what else can be interesting about this data.

**10 pts**

* 1. Show some plots that explain the task at hand, is it classification? or regression? Are you trying to predict a future event, or predict some additional information of a past event?
  2. state two self-supervised tasks as we have seen in class to pretrain your models for next sections on?

**35 pts**

1. Form a neural network graph based on the components we used in the walkthroughs in class.
   1. Decide your validation strategy for training your model   
      (subjects 107,108 are your test and should not be used during model training or evaluation of results)
   2. Create a naïve baseline solution and calculate train and validation score for that solution. This could be predicting last known value, or prediction of the class distribution for each category.
   3. Fit a classical machine learning model to the data and get a better and **solid** **benchmark** for the neural network model. Think which features will be useful for such a model
   4. Construct a neural network model and fit it to the data. analyze the results (Use visualizations to present your loss and other metrics you find relevant, show examples for good and bad classification with high probability, and refer to the uncertain predictions.   
      Compare the results you got on the training data vs. your results for the validation/test data)
   5. Pretrain your model on the one of the tasks you suggested in 1c and fine tune the trained model to the data. compare the results you got to previous sections (c-d)
   6. Try to figure out where & why is the model doing well and suggest at least 3 ways to improve the results
   7. Prioritize the list of suggestions for improvements and implement the first 2 suggestions, repeat section d. and present a table of your results (you may want to use a tool like Neptune.ml to record your results)

**Category embeddings**

**45 pts**

1. Using category embeddings within a network
   1. Join the [Favorita Grocery Sales Forecasting](http://www.kaggle.com/c/favorita-grocery-sales-forecasting) competition on kaggle.com, in this competition you are requested to predict the sales for each product and customer combination for the two weeks following the end of training data(Aug 15th-31st 2017) given the sales of the preceding ~4 years. Choose a group name that starts with “BGU-DL <group-name>” format
   2. Use a classical ML algorithm to predict the target and form a **solid** **benchmark**
   3. Perform the relevant preprocessing steps to create an embedding for the store id and for the item id (you may choose to create an embedding of additional dataset features that you think are relevant).
   4. Try predicting the target using only the features embeddings in a DL regressor, report the metrics you received on your training, validation and test set (submission to competition).
   5. Add the any other feature(s) you find relevant to your model and predict the target again using DL regressor (you may use any other type of network LSTM/GRU/CNNs in addition to model the time series part of the problem)
   6. Present any interesting insights that you got from the embeddings of categorical features
   7. Use the embeddings you got in 3d/3e as a “feature extractor” for a classical ML algorithm of your choice. How does your results for this combination compare to your previous results?

**10 pts**

1. Write a report that summarizes your research. Think what you would like to emphasize and what is of lesser importance present your analysis and report as a Kaggle kernel (you may train outside the kernels env. and use your saved model, if you prefer)

Your assignment submission should include a report to answer the above tasks (1-3), a screen capture of the place of your team in the private leaderboard, and a link to the Kaggle kernel that answers (4)