

Exercise 8 - deadline 21.05.2021 23:59

The dataset for this (and the following exercises) is the Penn World Table dataset.

You can find the latest version of the dataset here:

<https://www.rug.nl/ggdc/productivity/pwt/?lang=en>

Modify the exercise 7 to use time series. Compare the accuracy of the model with the one from exercise 7.

The goal for this exercise is to predict the national based real GDP (rgdpna) OR national based TFP (rtfpna) of each country based on other attributes **and the values from previous year(s)**. Be careful not to use false predictors!

Use a neural network for this purpose.

The goals of the exercise:

1. Select the target variable rgdpna or rtfpna.
2. Select the input features that you deem useful for the purpose.
3. Prepare the input attributes for the NN. (scaler)
4. Create the model and optimize the hyper-parameters.
5. Evaluate the final model.
6. Save the final model.

Optional goals:

7. Create a second script that loads the model, reads country data from country.csv and outputs the predictions in country_pred.csv
8. Create a third script that loads the model, and labeled data from country_feedback.csv. Using this feedback, print the accuracy of the NN on the feedback data. Then update the model using this data and save the updated model.

References:

https://scikit-learn.org/stable/modules/neural_networks_supervised.html#regression

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.html>

Important Note:

The submission should be done individually. Please name your script/notebook accordingly. At the beginning of your script / notebook give the name(s) of all students who worked on the exercise. You may work in small groups (of 2-3) but all students must be confident with the work of others. They should understand and be able to answer detailed questions of the implementation.