

Patrick Hutecker

02/12/2024

Professor Gu

Big Data Project Proposal

I will compare a handful of European data sets such as (Government Debt, Road Transportation, Food Supply, Tourism, Unemployment, GDP) to see if there is any correlation between data. The data will be gathered from *eurostat*.

I will first compare data sets by developing simple regression models – such as fitting the data to a linear transformation or looking at the covariance matrix. I will then see what fittings and covariances look most promising – and then investigate why these data sets might have an initial or naïve correlation. Then – if there are pragmatic reasons - rather than a correlation resulting from noise – I will apply a stronger – higher complexity fit – such as a Recurrence Neural Network – and use the size of the data and the gradient descent algorithm to fit the weights of the RNN.

The later part might be extensive – as the RNN would have to be well designed and justified – as well as coded with an advanced python library such as Tensorflow.

I foresee myself largely only using python or Google Collab to undergo this project – although it is important to note that documentation for the basic analysis will also have to be done in LaTeX.