Homework 2

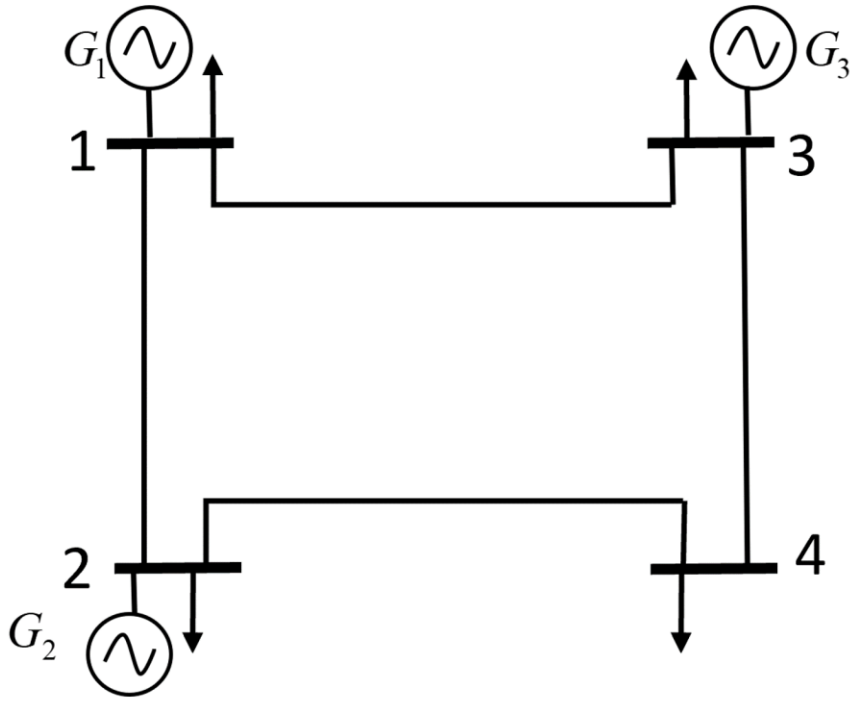
In recent researches, big data and machine learning are widely applied on the power system operation field. The power system operation data (e.g. power flow, bus voltage, generation data, loads' power demand, etc.) are enormous in terms of both dimensions and samples. With these data, system operators want to know the stability of a complex power system fast and accurately by the data-driven methods.

After learning this lecture on linear/nonlinear regression , please:

- Given the dataset 'hw2_3gpf.csv', apply the linear/nonlinear regression approaches to estimate the stability margin of the system.
- Try different regression methods to find the best fitting results. Try to understand the overfitting & underfitting phenomenon when you adjust the regression models and hyperparameters.

数据集描述

该文件中的前25列是某三机四线四节点系统（见4-bus test system.png或右图）的运行特征，分别是三台机组的有功出力、无功出力，四条线路的有功潮流、无功潮流，四个节点的电压幅值、电压相角，以及三台机组的理论有功上限。最后一列的Margin是基于运行状态计算得到的系统电压稳定裕度（可以认为是一个黑箱模型），每一行的sample之间互不相关。



请基于一定的特征选取和模型选择，合理划分训练集和测试集，对Margin的数值进行回归分析，并注意思考模型的过拟合、欠拟合问题。请提交两个作业任务对应的回归程序，.ipynb与.py文件均可。同时请附一份简要报告（pdf或word）对回归过程进行简要说明。