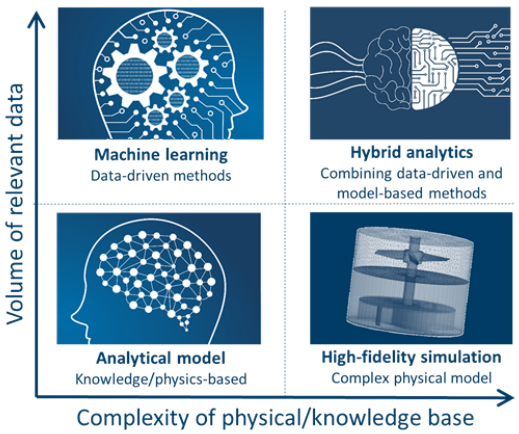
Robustness of Neural Networks

Background



The implementation of learning models in critical applications (e.g. autonomous vehicle transportation, data-based medical diagnosis, etc) is prohibited due the harmful consequences of erroneous model outputs. It has been difficult to pinpoint the exact cause for the lack of robustness in neural networks due to the combination of a large number of training parameters, nonlinearity of the model, non-convexity of the optimization problem used for training, etc [Nar and Sastry, 2019]. Recently the importance of persistency of excitation of the network weights has been proposed as a method to improve robustness of neural networks, and this project is to further investigate this direction.

This topic is proposed in connection with the project "Towards autonomy in process industries: Combining data-driven and model-based methods for estimation and control (TAPI)". The goal of TAPI is to move Norwegian land-based process industries towards more autonomous processes by exploring the potential merge of artificial intelligence (AI) and more traditional model-based control methods in order to understand, control and optimise the production processes. Project partners are NTNU, SINTEF, Hydro, Elkem, Yara and Borregaard.

Assignment

The goal of this assignment is to further investigate the importance of persistency of excitation in neural networks:

- Survey the literature looking at the connection between persistency of excitation and neural networks, and in particular make yourself familiar with [Nar and Sastry, 2019].
- Implement the method of [Nar and Sastry, 2019] to ensure persistency of excitation of all parameters of a neural network during training, and choose at least two test-cases for evaluating the benefit of their approach.

[Nar and Sastry, 2019] Nar, K & Sastry, S. Persistency of Excitation for Robustness of Neural Networks, arXiv:1911.01043, 2019

Main supervisor: Professor Tommy Gravdahl, tommy.gravdahl@ntnu.no

Co-supervisors: Esten Ingar Grøtli, EstenIngar.Grotli@sintef.no, Signe Moe Signe.Moe@sintef.no, Mark Haring Mark.Haring@sintef.no