0.1 Success Criteria

- Successfully reproduced Tishby's results as presented in his paper.
- Showed That Tishby's results are robust to changes in some hyperparameters

0.2 Extensions

- Extended the code to be able to reproduce results by Saxe.
- Reproduced some results Published by Saxe
- Extended the code to accommodate for different Datasets
- However, did not have the time to test if the results of Both Tishby and Saxe are robust to changes in datasets.
- Extended the code to be able to compare multiple MIE in real time on the same instance of NN.
- Implemented adaptive ways to skip MI computations in order to save on compute, Multi Threaded ones and Single Threaded ones, depending on how much memory is available.
- Implemented movie plotting which allows us to see training over time, and better analyse the data.

Tishby's Experiment Tishby[**?**] talks about a subset of the hyper parameters mentioned above.

Tishby varies the following

- Even the implementation of the mathematical ideas can be considered a hyper parameter – as it is unlikely that implementations follow the mathematical model exactly. This is the reason why independent verification is important and partly the reason of this thesis.
- Notes on comparing MIEs need to run on the same NN as they vary, need to run in real time as saving the whole network
- ¡++¿
- ¡++¿
- How to choose hyper parameters for binning, kde and batch. For the batching MIE, I have not investigated how we could choose a potential value of b such that the range parameters θ for epochs in the range$[e, ..., (e+b)]$ can be considered to be from the same probability distribution. And I am not aware of any method to choose the number of intervals in the Binning MIE or the noise size in KDE MIE.

0.3 Ending remarks

- Tishby and Saxe makes big claims about phases that a neural network is in NN. Such as "Is the compression phase an inherent part of NN and SGD algorithm" with out first agreeing on if compression can actually happen inside NNs.
- Furthermore, they are using very simple MIE.
- However they are using very simple MIE. In my opinion we need better tools before we can judge
- Success Criteria
- we have successfully reproduced the results showed by Tishby and Saxe. However there are reasons to not trust either of them as they have flaws with them.
  - Tishby – used only a toy dataset
  - Saxe – Changed allot of parameters at once made the claim that no compression phase happens
- We need better tools for MIE
  - cannot judge subtleties if something has a compression phase our MIE are not trustesd
  - we have seen KDE and Discrete show inconsistent results, when n'th layers hass less information about the input than the n+1'th layer.
- compression phase
  - judging if the networks have a compression phase is moot point as of yet as tools for measuring information are not good enough. Case and point Saxe argues that there cannot be compression in NN but every every experiments show that compression exists.
  - Saxe states that there is no compression in NN, however their experiments disagree.
  - we don't have the tools to say if the compression phase is actually happening
  - Tishby says compression phase happens but he is using a toy dataset, which was shown to not have a compression phase by Saxe.
- performance