

# Calculating Mutual Information in Deep Neural Networks

## Progress Report

Andrius Grabauskas, ag939  
Robinson College  
Thursday 31<sup>st</sup> January, 2019

**Project Supervisor:** Dr. Damon Wischik

**Overseers:** Dr. Robert Mullins

**Director of Studies:** Prof. Alan Mycroft

Prof. Pietro Lio'

### Success Criteria

Success criteria have been achieved, I am now able to reproduce the results in Tishby's paper<sup>1</sup>. That is, I am able to reproduce the information plane results that show the drift and the compression phase of a Neural Network for the specific parameters that Tishby used, and to show that the results are stable with variations to the parameters such as: batch size, network shape, training size.

### Current Progress, Planned work

I am not yet able to confirm or deny Tishby's results when varying the Dataset neuron activation function or the Mutual Information estimators (MIE). There has been some difficulty getting sensible results from the Mutual Information estimators that I was able to find, and I have reason to believe that Tishby's MIE might be flawed. The current plan in this regard is to try and use the same estimator that was used in Saxe's paper<sup>2</sup> which used a Kernel Density estimator (KDE) and binning to compute the Mutual Information (MI). Saxe's paper conducted an experiment where they changed Dataset, activation function and the MIE, my current aim is to only vary the MIE which would show if Tishby's result is only due to how he calculates MI.

Also future plans include exploring the idea of how we can consider weights between layers as random variables, and test if Tishby's hypothesis works under the those conditions. Tishby himself strongly suggests that weights should be considered as random variables but as of yet I am not aware of any work that tests this hypothesis. We are still unsure how it is best to test conduct such an experiment.

---

<sup>1</sup><https://arxiv.org/abs/1703.00810>

<sup>2</sup>[https://openreview.net/forum?id=ry\\_WPG-A-](https://openreview.net/forum?id=ry_WPG-A-)