Thesis proposal

February 2020

Brief project description with interim goals

Information-based models for deep learning has become a sought-after subject of ongoing research. Many recent methods for self-supervised and unsupervised representation learning train feature extractors using an estimate of the mutual information. Such an approach proved its ability to reach state-of-the-art performance in many applications. We believe that investigating the Information Bottleneck principle will give useful insights to find a key to understanding how to build more efficient and accurate algorithms to solve the tasks of deep learning.

The Information Bottleneck principle states that neural networks that discard irrelevant information from the input are generally more robust for predictive tasks.

With this project, we want to validate this theory by estimating the amount of predictive information $I(z;y)$ and input information $I(z;x)$ for neural networks trained with different regularization procedures (note that in this case z is the last layer activation).

First, we will define a metric to assess the generalization capabilities of a model. We plan to do this by measuring the amount of relevant information for representations, obtained to become an input to a classifier, and then evaluate the predictive capabilities of a model depending on the amount of data needed to build it. Then we will analyze how much $I(z;x)$ (and $I(z;y)$) for different trained models correlates with the previously defined generalization metric.

The possible investigation will involve one or more of the following:

- Vanilla neural networks
- Neural networks trained with different regularization strategies (such as Weight decay and dropout)
- Models trained with unsupervised objectives (VAEs, InfoMax, CPC)
- Bayesian Neural Networks
- Variational Information Bottleneck (VIB)
- A novel approach for VIB based on a different formulation of the same objective

The output of this project will result in an in-depth analysis of different popular models for at least 2 different datasets, one of which aims to show the validity in real-world scenarios. The analysis aims to investigate if the Information Bottleneck principle empirically correlates with the generalization capabilities of the different models.

Estimated timetable:

1) February-March: Project setup and literature exploration
1.1) Literature review
1.2) Definition of the benchmark tasks
1.3) Implementation and comparison of mutual information estimators
1.4) Implementation of the generalization evaluation metrics
1.5) Finalization of the standardized benchmarking framework

2) March-May: Comparison of the different architecture
2.1) Implementation of different neural network architectures
2.2) Experimental data collection phase
2.3) Analysis and visualization of the empirical results

3) June-August: Thesis writeup

Steps 2.1 to 2.3 will be repeated multiple times on different models. The total amount of iterations depends on time availability and relevant findings. The time required for step 1.3) might also vary depending on the variance registered on the empirical measurements as accurate Mutual Information estimation is an open problem in the research field.