MAT-VAE

CS726: Advanced Machine Learning

Group: ClosedAI

Advait Rahul, 200100014
Pulkit Paliwal, 20D100021
Sanchit Jindal, 200020120
Lakshya Gupta, 200050066

Project Report

Indian Institute of Technology, Bombay

Contents

1 Introduction

1.1 Representation Learning

Representation Learning or Feature Learning is an important subfield in Machine Learning which is concerned with learning representations of the data that make it easier to extract useful information when building classifiers or other predictors. In the case of probabilistic models, a good representation is the one that captures the posterior distribution of the underlying explanatory factors for the observed input. These learned representations play a crucial role in modern ML systems, serving various downstream tasks. Upon the advent of Deep Learning, several new techniques have been developed to learn representations from all sorts of data. For eg. Convolutional Neural Networks have been used to extract features from input images which are then used for various applications such as object recognition, edge detection. One such deep learning model that is used to learn representations is Variational Autoencoder which is the focus of this project.

1.2 Balancing Learning and Inference in VAEs

The classical Variational Autoencoders are trained using the Evidence Lower Bound (ELBO) objective. However, in practice, it has been found that the approximate inference distribution learnt by VAEs is often significantly different from the true posterior. Zhao et al. [2] claims that this problem is rooted in the ELBO objective.
Another problem pointed out by the authors is that when the conditional distribution is sufficiently expressive, the latent variables are often ignored. The ELBO objective encourages message $q_\phi(z \,|\, x)$ to be a random sample from $p(z)$ for each input x making the message uninformative about the input. Thus, if the decoder is very flexible, then a trivial strategy of only producing from $p(z)$ can globally maximize the ELBO objective leading to uninformatiove latent representations.
In order to overcome these limitations of VAE, Zhao et al. [2] proposed InfoVAE, a new family of Variational Autoencoders. Instead of the classical ELBO objective, the usage of a Maximum Mean Discrepancy (MMD) objective was shown to address the issues of uninformative latent code and variance over-estimation in feature space.

1.3 Obstacles in Representation Learning

Often times, when training these representations, it's common to encounter unknown computational and statistical constraints for each downstream task. For example, in a client-server setting, the server might have the necessary computational resources to handle data of 128 dimensions however, the client may not possess the same and can handle only 32-dimensional data. In such cases, we often restrict the capabilities of our models to meet the requirements of the user with the least resources. This results in under-utilization of resources for all other users in the system and a significant drop in the overall performance of the model. This issue can be resolved if the information present in the input data could be encoded at different granularities in a single embedding. However, the existing deep learning models tend to diffuse information across the entire representation vector due to their gradient-based training. In order to overcome this, Kusupati et al. [1] presented Matryoshka Representation Learning (MRL) which encodes

information at different granularities and allows a single embedding to adapt to the computational constraints of downstream tasks.
The modification allows for usage of smaller embedding sizes with a lower loss of performance as compared to training an individual model with a smaller embedding size. The authors demonstrated the performance of Matryoshka Representation for classification tasks by modifying the ResNet-50 model and comparing its performance on ImageNet-1k dataset.

2 Problem Formulation

This project seeks to integrate Matryoshka Representation Learning and InfoVAEs to create a VAE model capable of delivering quality flexible representations. With the help of MRL, we intend to overcome the need of training multiple low-dimensional models and instead have a singular model which produces a single embedding that can be deployed adaptively at no additional cost during inference.
We showcase the effectiveness of MRL-MMD-VAEs by evaluating their performance against individually trained MMD-VAEs and standard VAEs in a basic classification task using the MNIST and Fashion-MNIST datasets.
For implementing these models, we'll primarily refer to the following Github repositories

1. InfoVAE

2. MRL

References

[1] Kusupati, A., Bhatt, G., Rege, A., Wallingford, M., Sinha, A., Ramanujan, V., Howard-Snyder, W., Chen, K., Kakade, S., Jain, P. and Farhadi, A., 2022. Matryoshka representation learning. Advances in Neural Information Processing Systems, 35, pp.30233-30249.

[2] Zhao, S., Song, J. and Ermon, S., 2017. Infovae: Information maximizing variational autoencoders. arXiv preprint arXiv:1706.02262.