Benchmarking the hyper-parameter sensitivity of VAE models for cancer treatment

Armin Korkic A.Korkic@student.tudelft.nl

1. Introduction

- Improve cancer treatment through Machine Learning.
- However, these algorithms are black-boxes.
- In this project we try to take these black curtains away.
- By benchmarking these models we try to measure their sensitivity to hyper-parameters changes.

2. Research Question

"How sensitive are different VAE models to the choice of hyper-parameters ?"

3. Method

Performing a Grid Search over the following hyper-parameters

- Learning rate = [0.1, 0.01, 1e-3, 1e-4, 1e-5]
- Latent space = [10, 20, 30, 50, 100, 200]
- Optimizers = [Adam, SGD, RSMprop]

VAEs on which a grid search is performed:

- VAE, IWAE, Info-VAE, LogCosh-VAE

The Data

- The data used for the benchmarking stems from The Cancer-Genome Atlas [1].
- The data represents the RNA Gene Expression of different cancer tissues.
- Original data contains 11000 features, but for this experiment only the 5000 most variable ones are used.

4. The Models

VAE

- A Variational Auto-Encoder (VAE) is a type of neural network which is derived from Auto-Encoders.
- A VAE functions similarly as a Auto-Encoder, but instead of compressing the data into a single point in the latent space. It creates a distribution of the latent space.

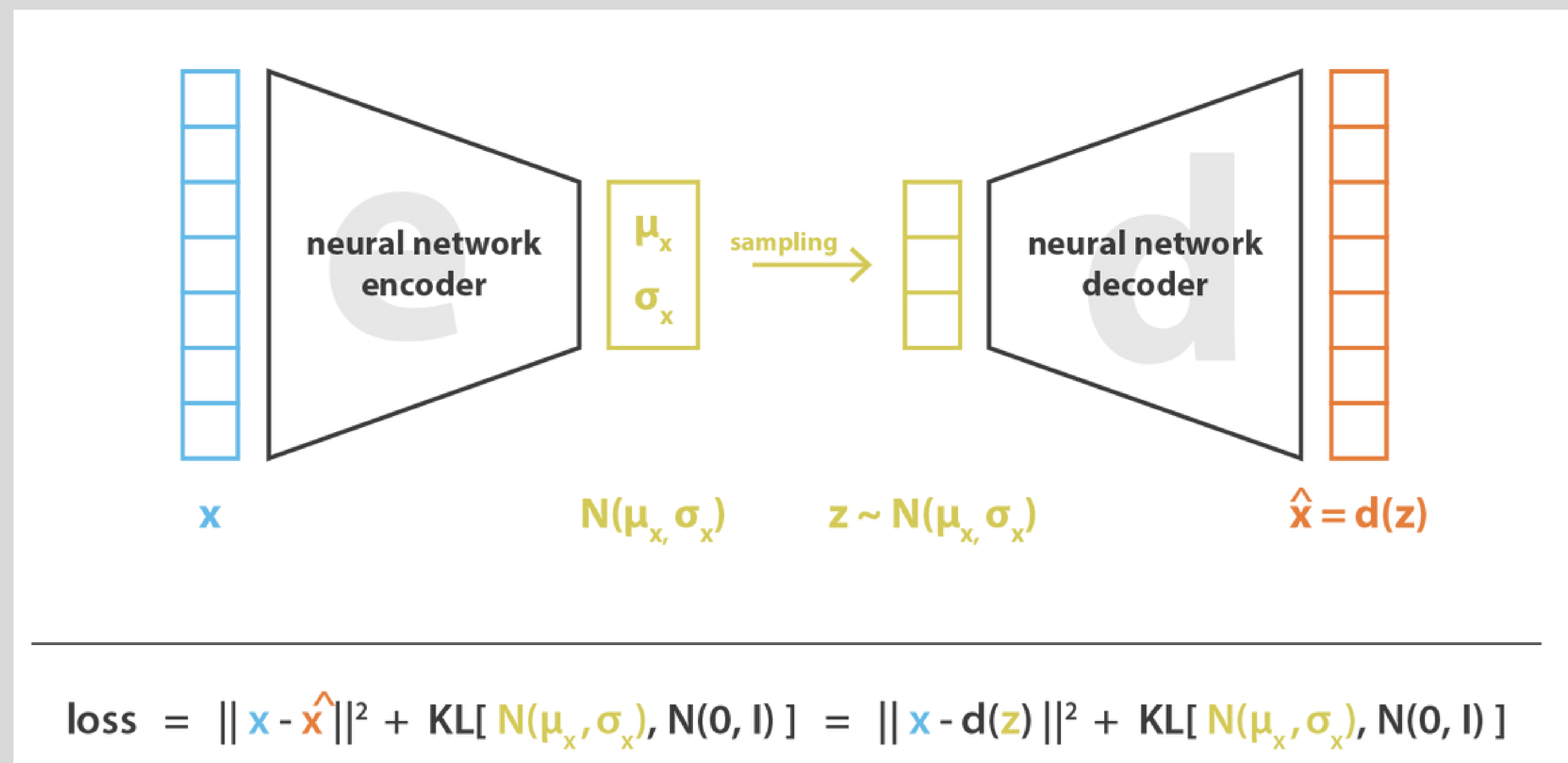
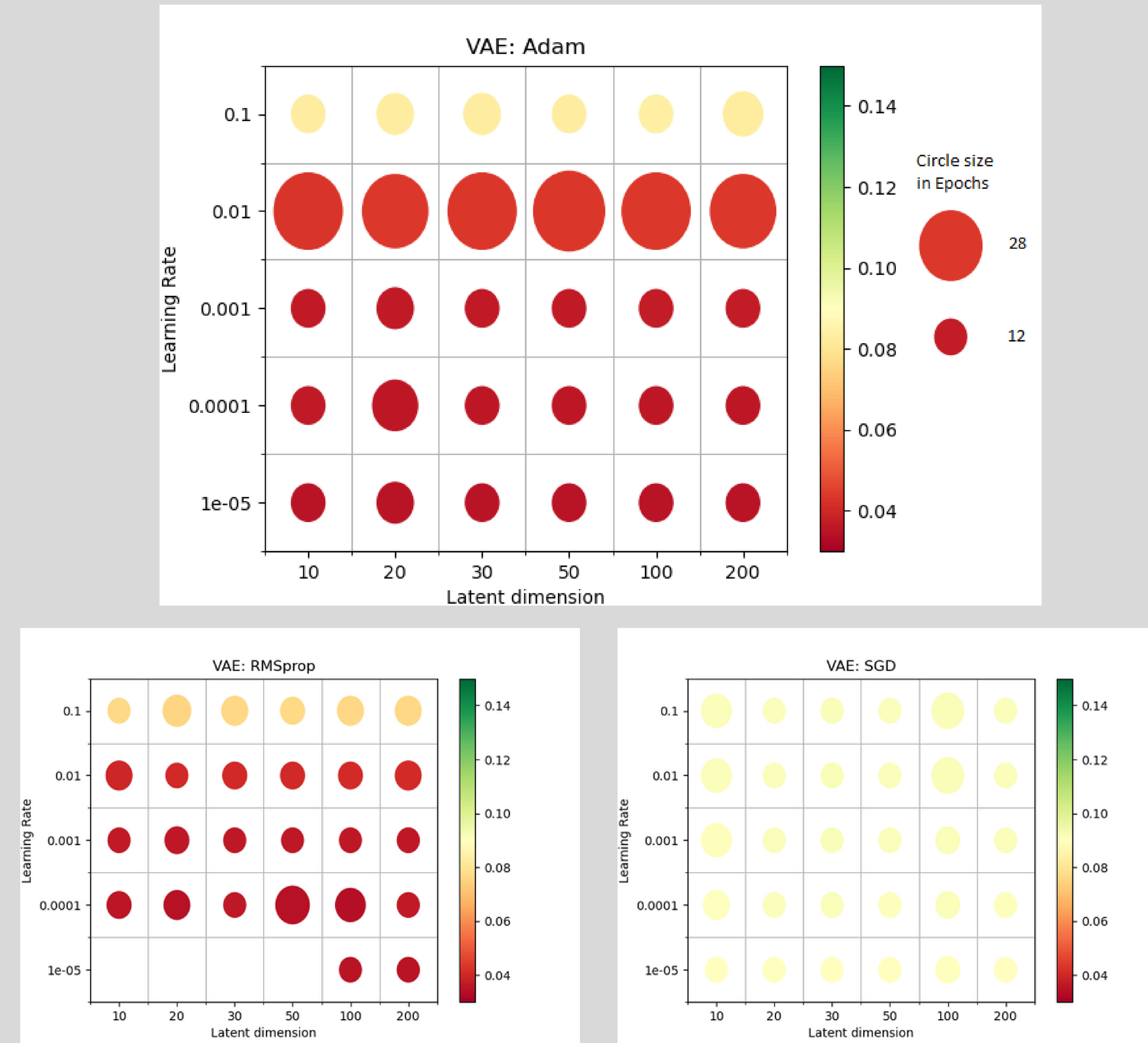


Figure of a VAE, with its loss function retrieved from; https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Differences between these models relative to the Vanilla-VAE

- **IWAE**; Introduces importance sampling . Takes into account multiple samples to approximate posterior [2].
- **Info-VAE**; It replaces ELBO (Evidence LowerBound) criteron for MMD (Maximum Mean Discrepancy). It tries to solve the problems of "Uninformative Latent Code" and "Variance Over-estimation in Feature Space" [3].
- **LogCosh-VAE**; Replaces the Mean Square Error from the standard VAE with the hyperbolic cosine loss [4].

5. Results & Conclusion



The x-axis represents the Latent dimension, the y-axis the learning rate. The color and the size of the circles, are the reconstruction loss and the amount of epochs untill convergence respectively. The largest circles amount to 28 epochs and the smallest to 12 epochs.

- The optimizer plays the most important role performance wise.
- Models are much more sensitive to changes to their optimizer.
- The best optimizers are Adam and RMSprop.
- The learning rate is also important when configurating the models.
- Most models perform the best with a learning rate of 0.001.
- Some configurations result in NaN values as output. Happens in certain configurations with extreme values s.a. high learning rates in combination with low latent dimensions.

References:
[1] UCSC Xena. (2016, December 29). Gene Expression RNA seq.
[2] Burda, Yuri \& Grosse, Roger \& Salakhutdinov, Ruslan. (2015). Importance Weighted Autoencoders.
[3] Zhao, Shengjia, A Tutorial on Information Maximizing Variational Autoencoders (InfoVAE).
[4] Chen, Pengfei & Chen, Guangyong & Zhang, Shengyu. (2018). Log Hyperbolic Cosine Loss Improves Variational Auto-Encoder

Responsible Professor: Marcel Reinders
Supervisors: Stavros Makrodimitris, Tamim Abdelaal, Mohammed Charrout, Mostafa elTager
Date: 02-07-2021