

Exercise 4

*Nir Raviv 200683548**Roi Tabach 203022983**Andrey Leshenko 322026527**nirraviv@mail.tau.ac.il**roi.tabach@gmail.com**andrey.leshenko@gmail.com*

Q1

The code implementing VAE is on NOVA at [~leshchenko/arazim/advanced_ml/public/](#). To run it, you must first **enable a local Python environment**, by running:

```
$ source ./keras_env/bin/activate.csh
```

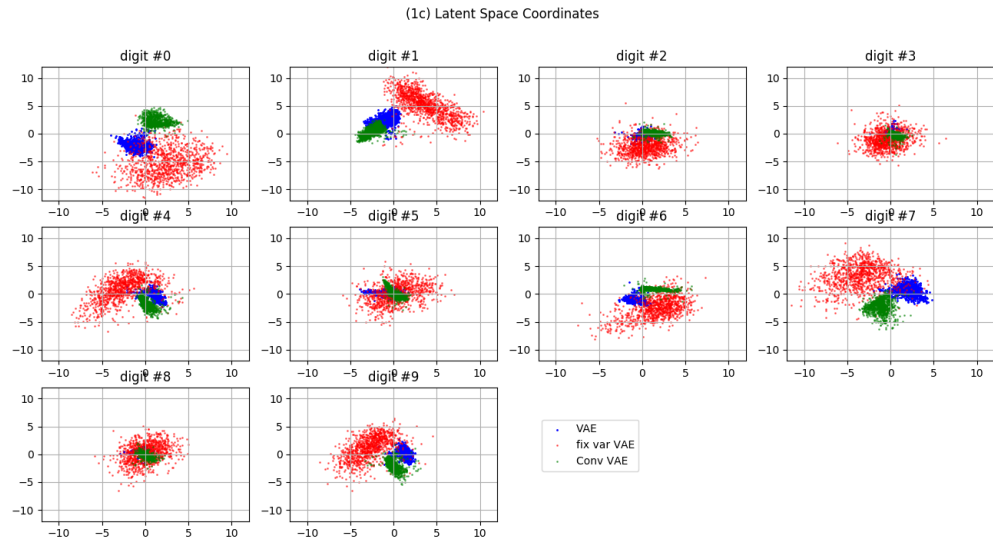
Now if you run a script using “python something.py”, this environment will be used. You can run the trained network in the [hw4_trained](#) folder in the same directory, or first train and then run in [hw4_untrained](#). If you start a long running task on NOVA, it will usually be killed by the server after a few minutes. For this reason, if you want to train through an SSH connection to NOVA, first SSH to one of the Shreiber PCs (for example gauss-11) and train there.

Important note about plots: by default the code will display the plots in a window. If when connecting to NOVA you passed the -X argument to SSH, you should be able to see these windows. If opening windows is not possible and you would prefer to save the plots as image files, open the Python code and change **SAVE_PLOTS** to True.

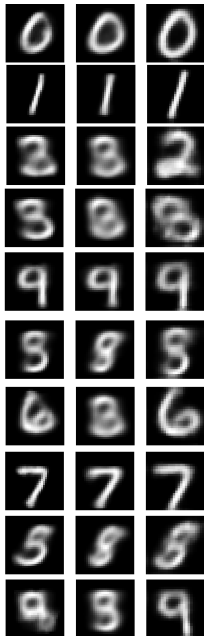
The following sections, c,d,e, will be answered for all models: VAE, fixed-variance VAE, and Convolutional VAE. The answers for f,g are going to be included in c-e.

c

First we will show the latent-space coordinates of the entire MNIST-dataset, for each of the representations.



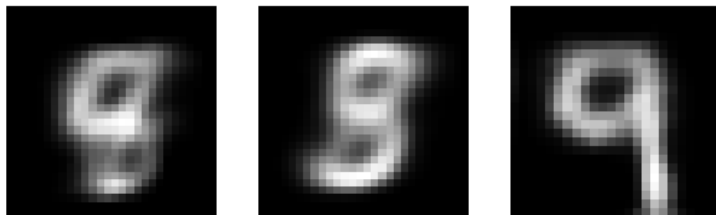
As one can see the VAE and Convolutional VAE latent space points have small overlap since their variance is smaller. Therefore, these auto encoders are better. Next we show an example for each of the digits, 0-9 (each in it's seperate row), the representation reconstructed with each of the decoders: left is the VAE, middle is the fixed variance VAE, and right is convolutional VAE.



d

Here we take decode the latent-space vector $(0.5, 0.2)$ using each of our decoders: left is the VAE, middle is the fix-var VAE, and rightmost is the conv-VAE.

$(0.5, 0.2)$ decoded with VAE, fix-var VAE, conv-VAE



e

Here we show how interpolation between two latent-space points look in the decoded space, using each of our decoders: VAE:



Fixed Variance:



Convolutional:



Q2

Given $P \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Q \sim \mathcal{N}(\mu_2, \sigma_2^2)$ random variables, we need to derive an expression for the KL divergence of P and Q .

$$\begin{aligned}
 D_{KL}(P||Q) &= \int_{\mathbb{R}} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx = \mathbb{E}_p[\log(p(x))] - \mathbb{E}_p[\log(q(x))] \\
 \mathbb{E}_p[\log(p(x))] &= \mathbb{E}_p \left[\log \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \right) \right] = -\frac{1}{2} \log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2} \mathbb{E}_p[(x - \mu_1)^2] = \\
 &= -\frac{1}{2} \log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2} \mathbb{E}_p[x^2 - 2x\mu_1 + \mu_1^2] = -\frac{1}{2} \log(2\pi\sigma_1^2) - \frac{\sigma_1^2}{2\sigma_1^2} = -\frac{1}{2} \log(2\pi\sigma_1^2) - \frac{1}{2} \\
 \mathbb{E}_p[\log(q(x))] &= \mathbb{E}_p \left[\log \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \right) \right] = -\frac{1}{2} \log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2} \mathbb{E}_p[(x - \mu_2)^2] = \\
 &= -\frac{1}{2} \log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2} \mathbb{E}_p[x^2 - 2x\mu_2 + \mu_2^2] = -\frac{1}{2} \log(2\pi\sigma_2^2) - \frac{\sigma_1^2 + \mu_1^2 - 2\mu_1\mu_2 + \mu_2^2}{2\sigma_2^2} = \\
 &= -\frac{1}{2} \log(2\pi\sigma_2^2) - \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2}
 \end{aligned}$$

Thus,

$$D_{KL}(P||Q) = -\frac{1}{2} \log(2\pi\sigma_1^2) - \frac{1}{2} + \frac{1}{2} \log(2\pi\sigma_2^2) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} = \log \left(\frac{\sigma_2}{\sigma_1} \right) - \frac{1}{2} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2}$$