Setup

I created a dataset of three different shapes:
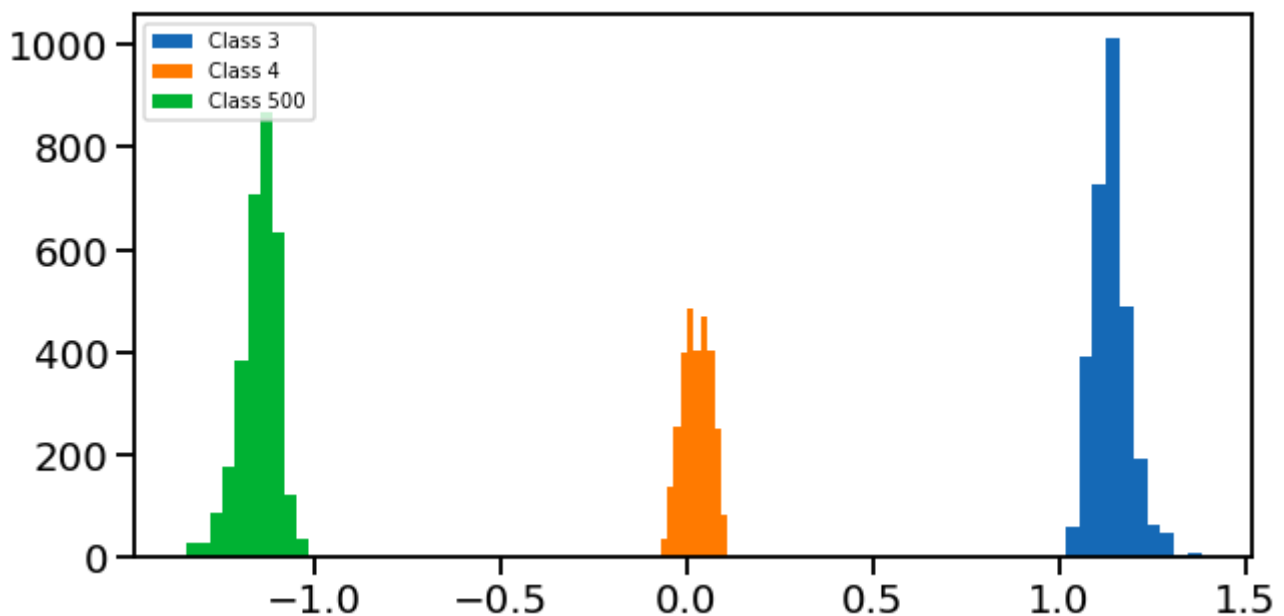
1. Triangle
2. Square
3. Circle (approximated as a polygon with n=500 sides)

It had 10,000 examples in total and every shape was (roughly) equally represented.

1. On Clean Data

This dataset is *without* any kind of noise. The only variation is in shape, and within shapes there is no variation. For example, all triangles are the same.

Then I trained Vanilla VAE with 1 unit in the latent layer. Following is the plot of latent unit activation (mean of our learned latent dist).

As can be seen, there is very clear separation of three classes.

And here are the reconstructions. Recons are generally very good and sharp, but some cheating behaviour (phantom square behing triangle) can be seen as well.
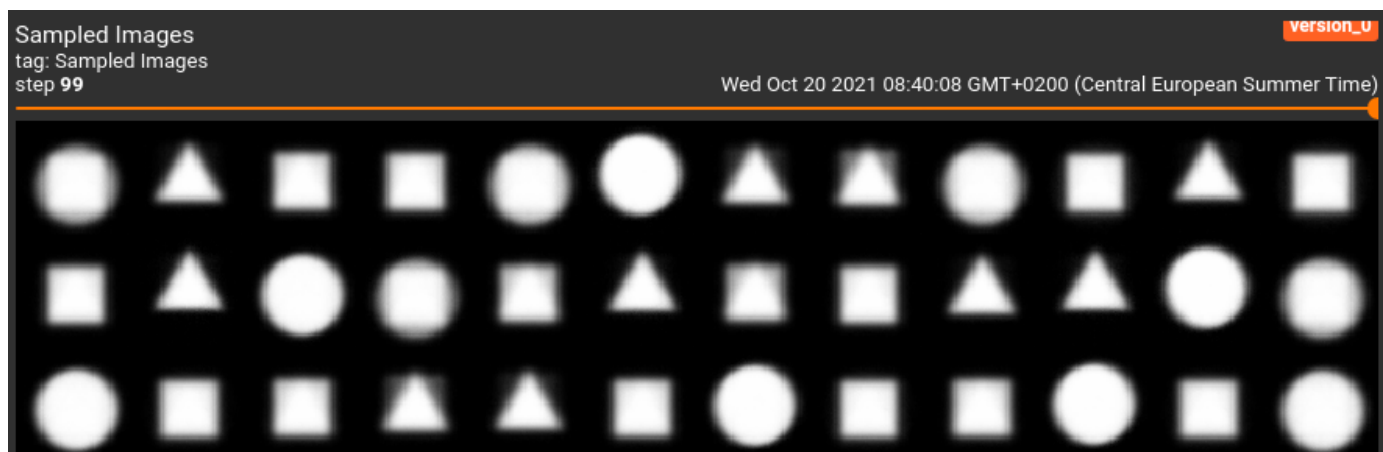
2. On Perturbed Data

Then I repeated this experiment, but this time I added noise. The noise is added to the position of shape's center i.e. it offsets the square from center point by a disturbance d ~ N(3,1)

Training and visualizing the means give us the following plot. This time the distributions are not as sharp, but they're still relatively compact. Square dist has been squeezed b/w circle and triangle dist.
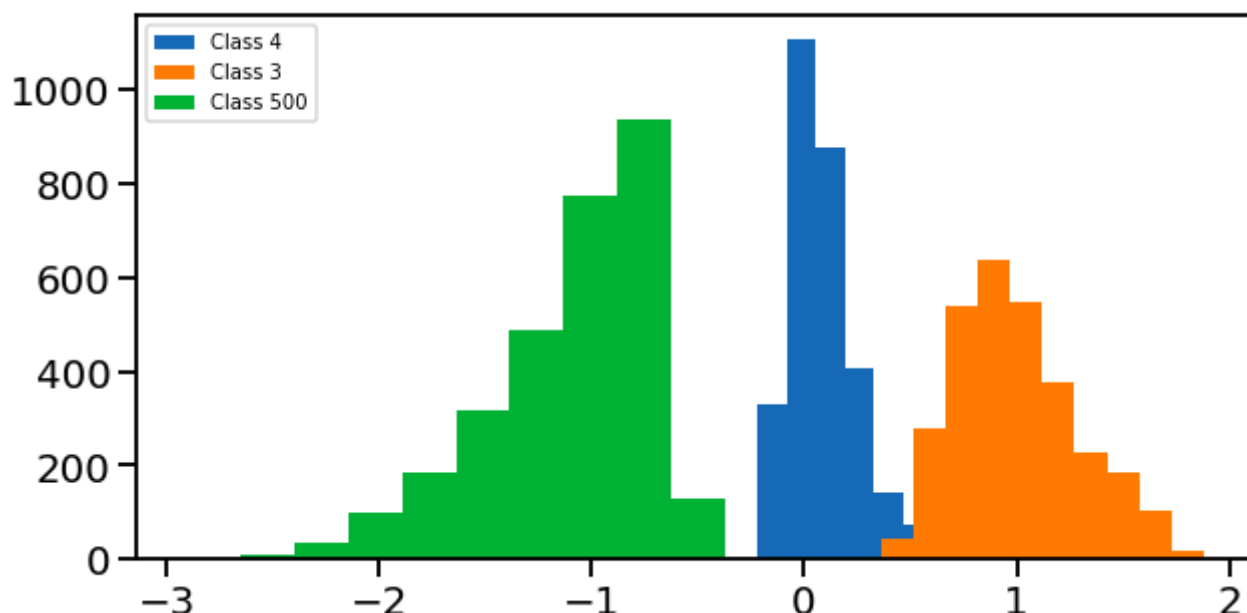
And here are the reconstructions on noisy dataset. Same cheating behaviour as before, but now the shapes are blurred as well.


3. On Perturbed Data with 2 Latents

At this point I thought that network is being forced to use the same dist to represent center position as well (as it is no longer fixed because of noise). This could explain higher variance in resultant dists.

To test this, I added one more latent unit with the assumption that network will try to capture the noise of center position in the second unit and keep first unit free of any interference due to noise.
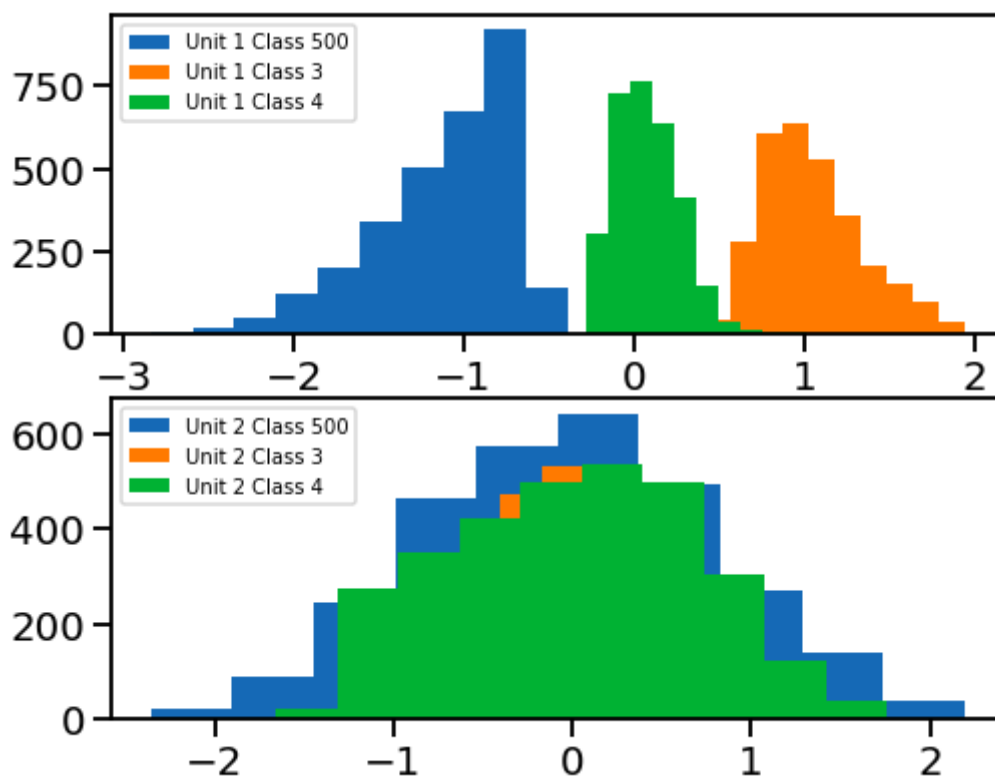
Since we now have 2 latent units, we can make a 2-D scatter plot of representations. As shown:
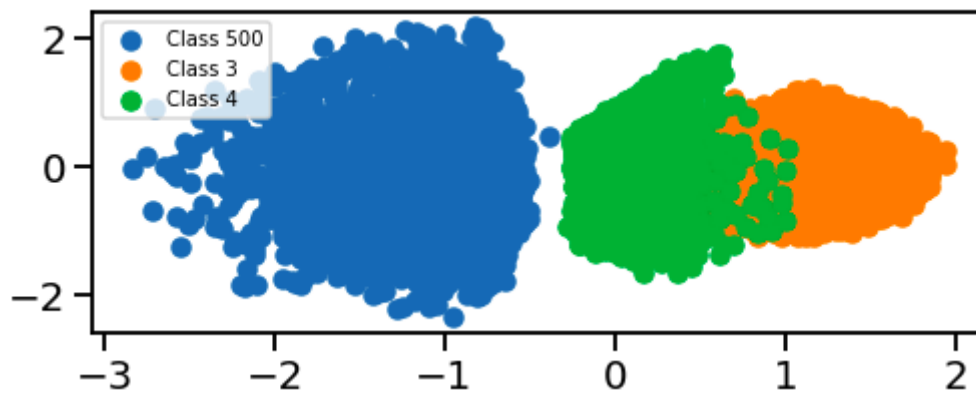
The distributions are still compact, and we again see square dist being squeezed b/w circle and triangle.

## Separate Activation Histograms for each Unit

I also plotted the histogram of each unit separately to see how they differ.

- Unit 1's histogram is the same as last time. We see more spread compared to clean data case.
- Unit 2's histogram shows that distribution is same in all three classes. This is not surprising since the noise we added was same for each class. What *is* surprising is that this didn't have any effect on the spread of unit 1's distributions

Reconstructions

And here are the recons when the network uses 2 units:
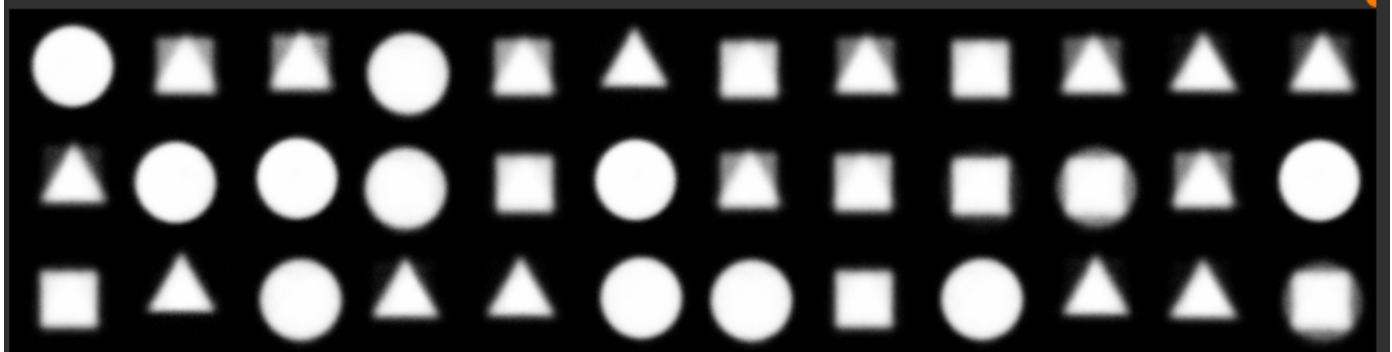
Recons look subjectively the same as in 1 unit case, perhaps slightly less blurred and network achieves a slightly lower loss value
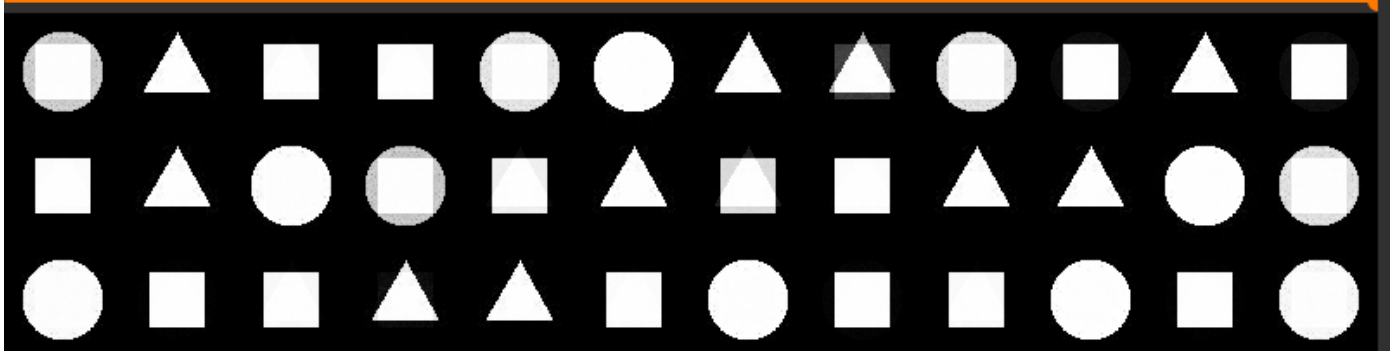
Type	Loss
Clean data (1 unit)	0.0217
Noisy data (1 unit)	0.06914
Noisy data (2 units)	0.06613

Some Interesting things to note:

1. Both histograms are limited to the [-2,2] range, even though this range is not being enforced anywhere. We do enforce a similarity to N(0,1) via KL-divergence.
2. The distribution corresponding to circle is the most spread out. This probably has to do with the fact that circle has the highest area among the three chosen shapes. This suggest that I should try out shapes with same (unit) area. I didn't enforce 'same area' constraint this time.
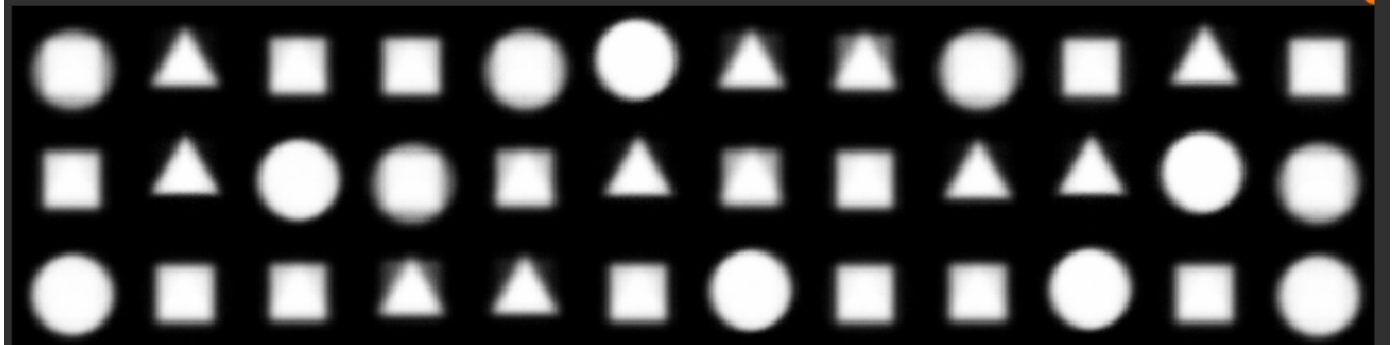
Recon comparison

clean data

noisy data (1 unit)

noisy data (2 units)

In [ ]: