

# VAEs in manufacturing

2018-05-25

---

dataprophet<sup>TM</sup>  
*machine learning specialists*



# Steel production

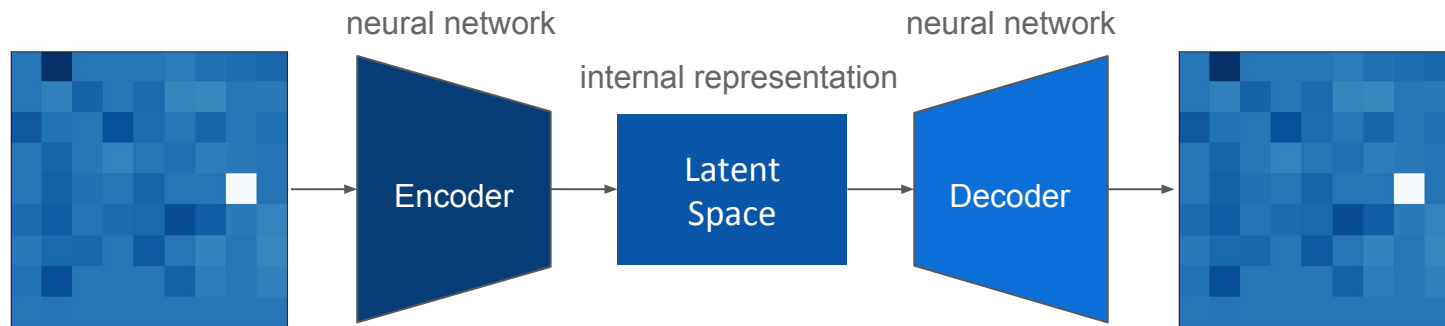


Billet Count Sheet

"A / B Run-out Billet Count Sheet"																	
Heat #		Grade		Crew		Total Std 1 / 4		Total Std 2 / 5		Total Std 3		Total Std 4		Total Std 5		Total Std 6	
Date		Heavy or Light		Run-out Operator		A		B		C		D		E		F	
BB SIZE:		Flange Gauge Inspection prior to mill performed by:		BB SIZE:		BB SIZE:		BB SIZE:		BB SIZE:		BB SIZE:		BB SIZE:		BB SIZE:	
CUT		Feet		Actual Wt.		Set point Wt.		CUT		Feet		Actual Wt.		Set point Wt.		CUT	
1	5.5	10.11	10.11	1	5.5	10.11	10.11	1	5.5	10.11	10.11	1	5.5	10.11	10.11	1	5.5
2	5.5	10.11	10.11	2	5.5	10.11	10.11	2	5.5	10.11	10.11	2	5.5	10.11	10.11	2	5.5
3	5.5	10.11	10.11	3	5.5	10.11	10.11	3	5.5	10.11	10.11	3	5.5	10.11	10.11	3	5.5
4	5.5	10.11	10.11	4	5.5	10.11	10.11	4	5.5	10.11	10.11	4	5.5	10.11	10.11	4	5.5
5	5.5	10.11	10.11	5	5.5	10.11	10.11	5	5.5	10.11	10.11	5	5.5	10.11	10.11	5	5.5
6	5.5	10.11	10.11	6	5.5	10.11	10.11	6	5.5	10.11	10.11	6	5.5	10.11	10.11	6	5.5
7	5.5	10.11	10.11	7	5.5	10.11	10.11	7	5.5	10.11	10.11	7	5.5	10.11	10.11	7	5.5
8	5.5	10.11	10.11	8	5.5	10.11	10.11	8	5.5	10.11	10.11	8	5.5	10.11	10.11	8	5.5
9	5.5	10.11	10.11	9	5.5	10.11	10.11	9	5.5	10.11	10.11	9	5.5	10.11	10.11	9	5.5
10	5.5	10.11	10.11	10	5.5	10.11	10.11	10	5.5	10.11	10.11	10	5.5	10.11	10.11	10	5.5
11	5.5	10.11	10.11	11	5.5	10.11	10.11	11	5.5	10.11	10.11	11	5.5	10.11	10.11	11	5.5
12	5.5	10.11	10.11	12	5.5	10.11	10.11	12	5.5	10.11	10.11	12	5.5	10.11	10.11	12	5.5
13	5.5	10.11	10.11	13	5.5	10.11	10.11	13	5.5	10.11	10.11	13	5.5	10.11	10.11	13	5.5
14	5.5	10.11	10.11	14	5.5	10.11	10.11	14	5.5	10.11	10.11	14	5.5	10.11	10.11	14	5.5
15	5.5	10.11	10.11	15	5.5	10.11	10.11	15	5.5	10.11	10.11	15	5.5	10.11	10.11	15	5.5
16	5.5	10.11	10.11	16	5.5	10.11	10.11	16	5.5	10.11	10.11	16	5.5	10.11	10.11	16	5.5
17	5.5	10.11	10.11	17	5.5	10.11	10.11	17	5.5	10.11	10.11	17	5.5	10.11	10.11	17	5.5
18	5.5	10.11	10.11	18	5.5	10.11	10.11	18	5.5	10.11	10.11	18	5.5	10.11	10.11	18	5.5
Flange size:		Flange size:		Flange size:		Flange size:		Flange size:		Flange size:		Flange size:		Flange size:		Flange size:	
Cut #	Top West	Bot. West	Top East	Bot. East	Cut #	Top West	Bot. West	Top East	Bot. East	Cut #	Top West	Bot. West	Top East	Bot. East	Cut #	Top West	Bot. West
2	1/4	1/4	1/4	1/4	2	1/4	1/4	1/4	1/4	2	1/4	1/4	1/4	1/4	2	1/4	1/4
4	1/4	1/4	1/4	1/4	4	1/4	1/4	1/4	1/4	4	1/4	1/4	1/4	1/4	4	1/4	1/4
6	1/4	1/4	1/4	1/4	6	1/4	1/4	1/4	1/4	6	1/4	1/4	1/4	1/4	6	1/4	1/4
8	1/4	1/4	1/4	1/4	8	1/4	1/4	1/4	1/4	8	1/4	1/4	1/4	1/4	8	1/4	1/4
Weights		Weights		Weights		Weights		Weights		Weights		Weights		Weights		Weights	
#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11
#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11
#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11
#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11	#BB	10.11
Scrap Ft.		Scrap Ft.		Scrap Ft.		Scrap Ft.		Scrap Ft.		Scrap Ft.		Scrap Ft.		Scrap Ft.		Scrap Ft.	
COMMENT:																	

- Steel producer
- Massive I-beams are cast and then milled into various shapes to be shipped to their clients.
- Variations in the shape of the I-beams can cause milling defects.
- Quality measurements are recorded manually.
- How much noise in billet quality measurements can we tolerate?

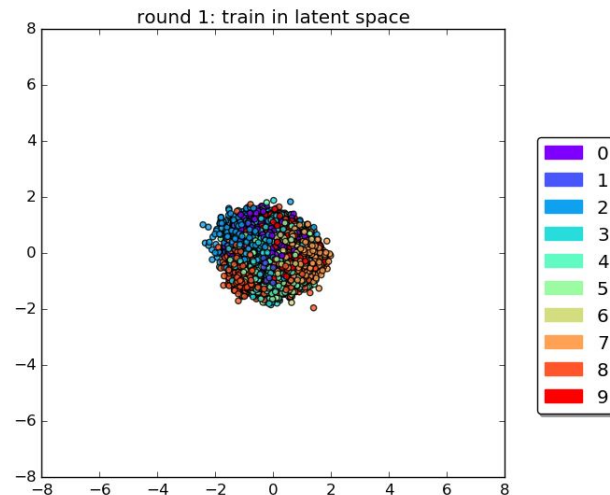
# Variational Autoencoders - an introduction



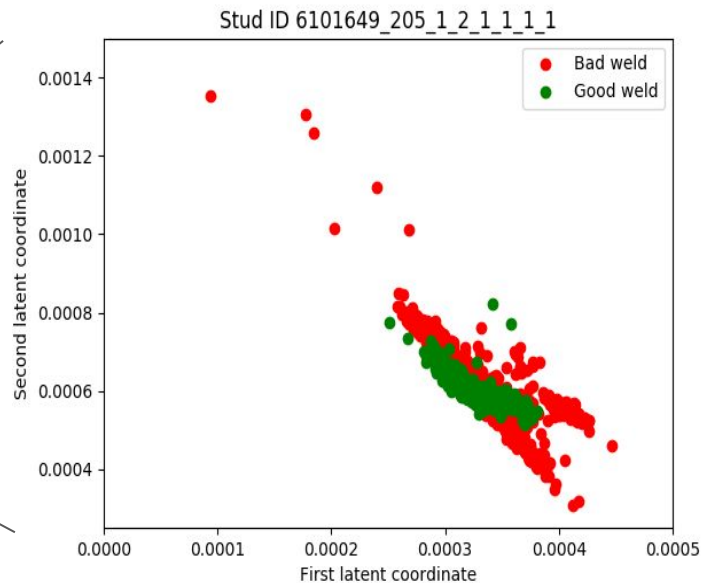
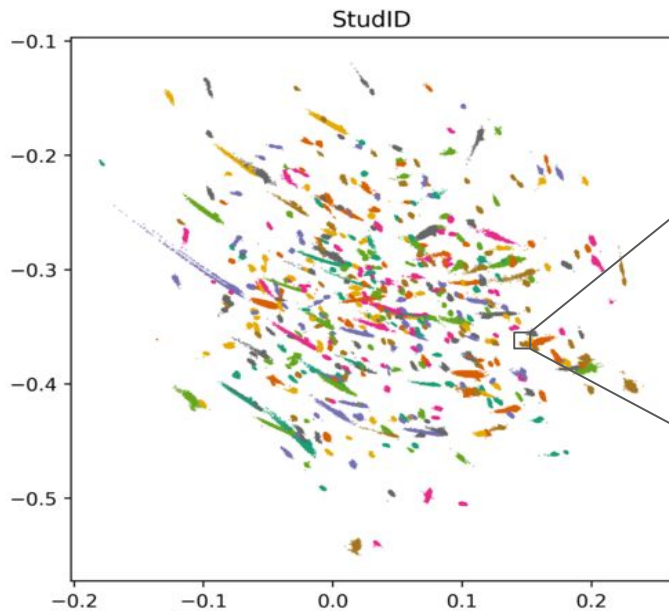
**Training dynamic:** Looking at the 2-dimensional latent space as the model is trained. Each sample is coloured by the type of sample it represents.

It can be seen that the model learns to separate different types of samples, but clusters the same type of sample together.

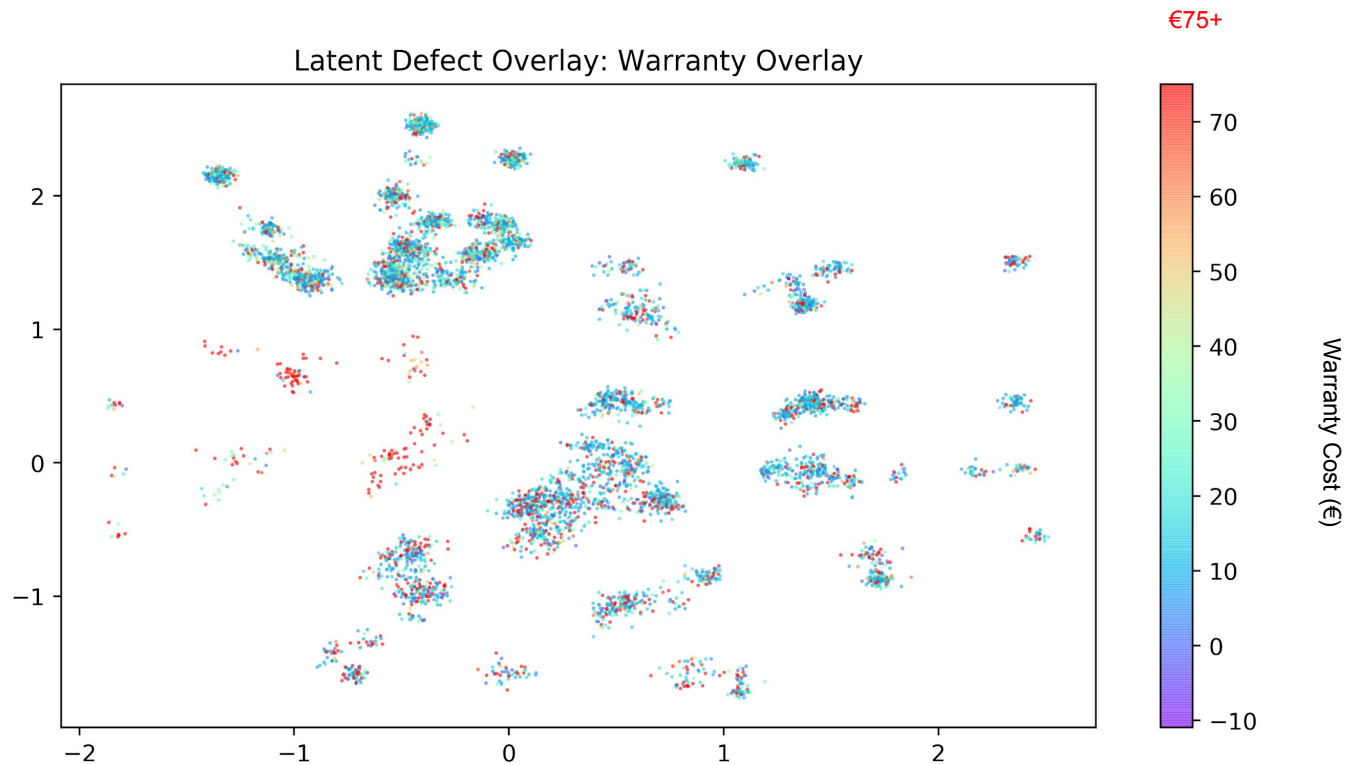
**The latent space encodes the internal representation, with similar samples clustered together.** The latent space does not have any directly interpretable/intrinsic meaning. Once the latent dimensions are decoded the rich embedding is discovered.



# Modelling: Latent Interrogation



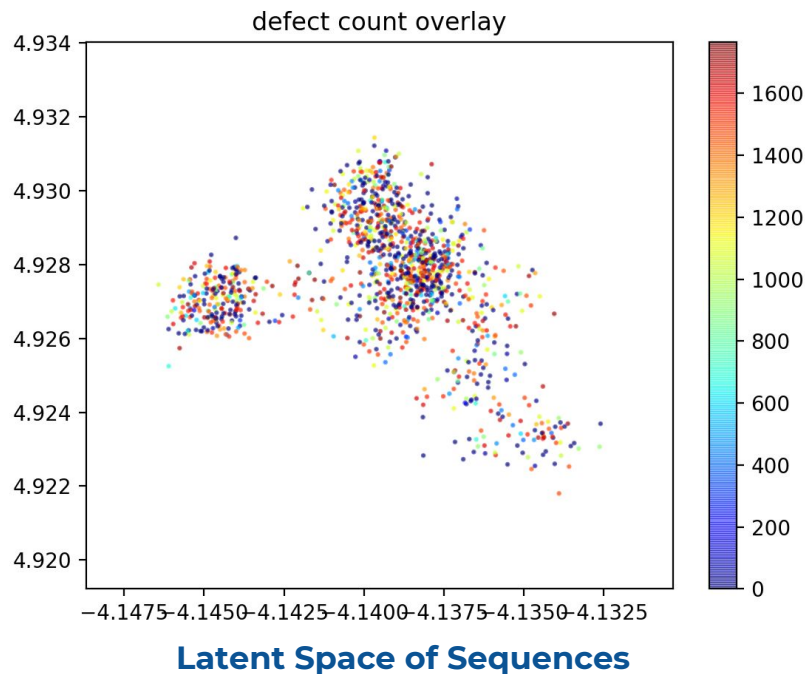
# Modelling: Latent Interrogation

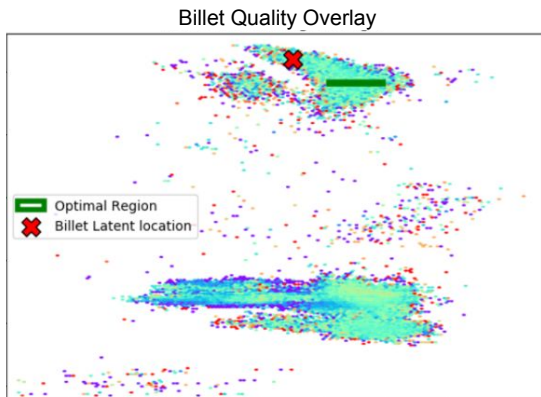


# Latent Space with Defect Count Overlay

**Input:** Actual production sequences of cars, with daily aggregated defect count as overlay.

**Result:** 2 cluster groups. There is no particular structure or relationship between the clusters and defect count.





Our application of VAEs is novel and largely unstudied in ML literature. We seek to understand the theoretical capabilities and/or limitations of the approach.

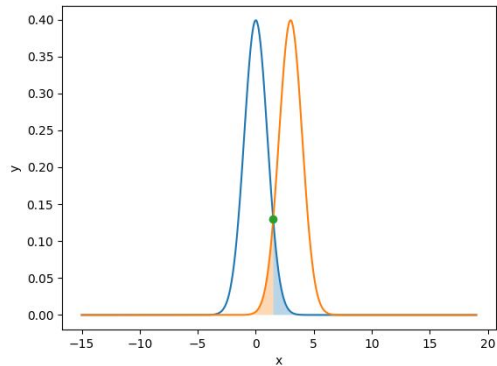
## Things we don't know:

- **How much latent separation can we expect, given an arbitrary high-dimensional dataset?**
- How can we isolate quality regions in a latent representation where complete separation was not achieved?
- How much noise can we tolerate in the quality measurement before results can no longer be trusted?
- What are the effects of changing the weighting of the KL-divergence and reconstruction terms in the loss function?
- *Other questions we haven't thought of asking yet...*



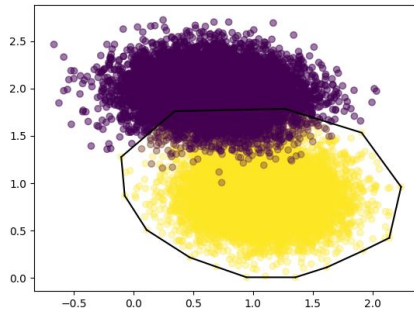
- Input data contains ten independent normally-distributed features.
- Two distinct groups ( $g_1$  and  $g_2$ ) exist in the data.
  - Different groups are sampled from two distinct and independent multivariate Gaussian distributions,  $f_1$  and  $f_2$ .
  - Both distributions have the identity covariance matrix.
- The overlap coefficient (OVL) is the area of overlap between the two distributions.

Data Distributions



- The purity ( $\rho_g$ ) is the proportion of all data points inside the convex hull around group  $g$  that belong to group  $g$ .
- For two equally-sized populations, the minimum possible purity for either group is 0.5.
- Latent separation is the harmonic mean of  $\rho_{g1}$  and  $\rho_{g2}$ .

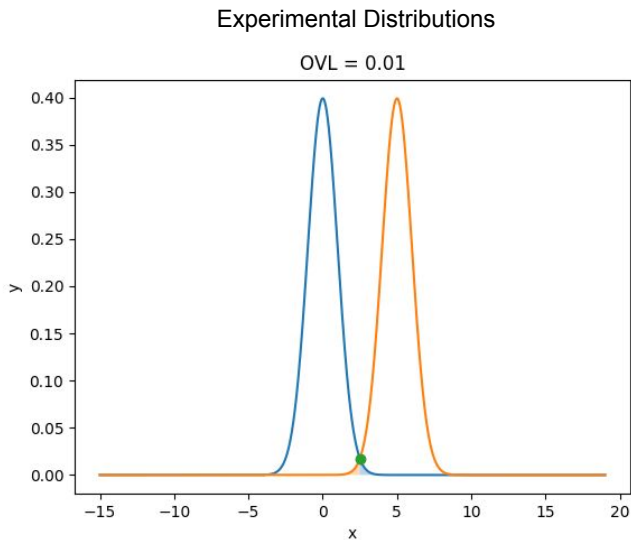
Latent Separation







- Initialise two 10-dimensional Gaussian distributions with univariate distributions for the individual features:
  - $\mu_1 = 0, \mu_2 = 5$ .
  - $\sigma_1 = \sigma_2 = 1$ .
- Decrease  $\mu_2$  in steps to systematically increase *OVL*:
  - Nine steps of size 0.5 for a total of ten *OVL*s.
- For each *OVL*, generate ten independent datasets:
  - Each dataset contains 20,000 samples (10,000 samples from each of the two distributions).
- For each dataset, train a VAE and measure the separation in the latent space:
  - Xavier weight initialization.
  - 10 training epochs.





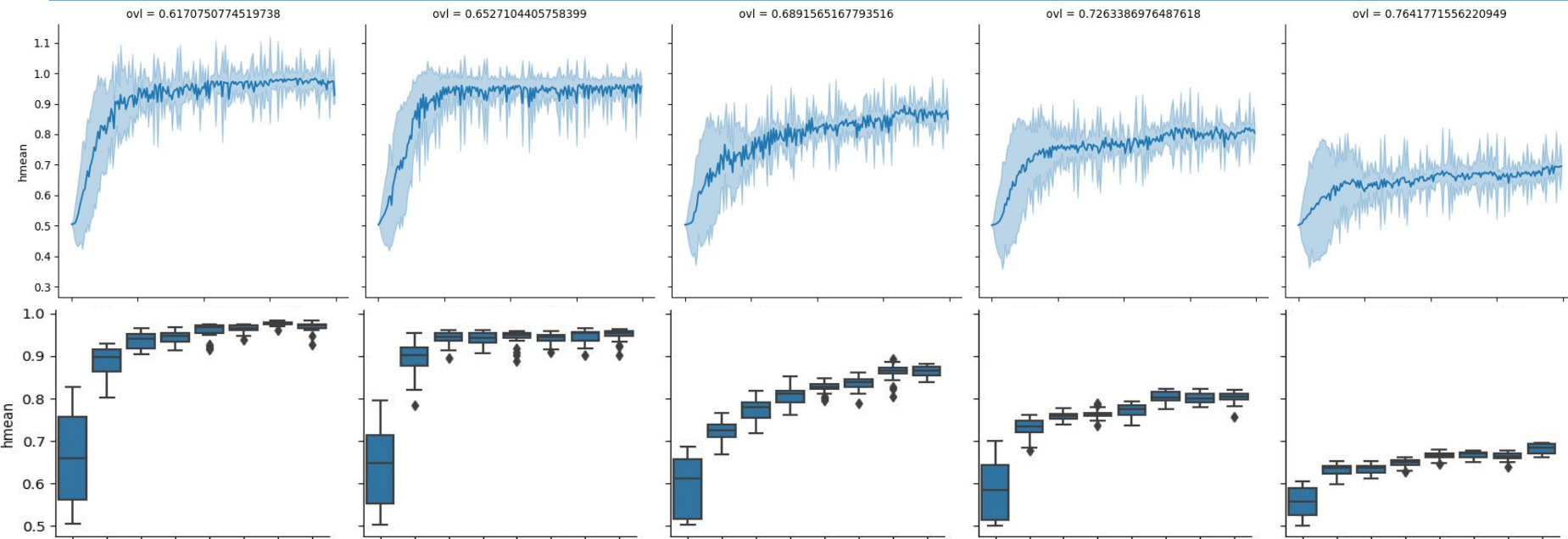
# Experimental Setup

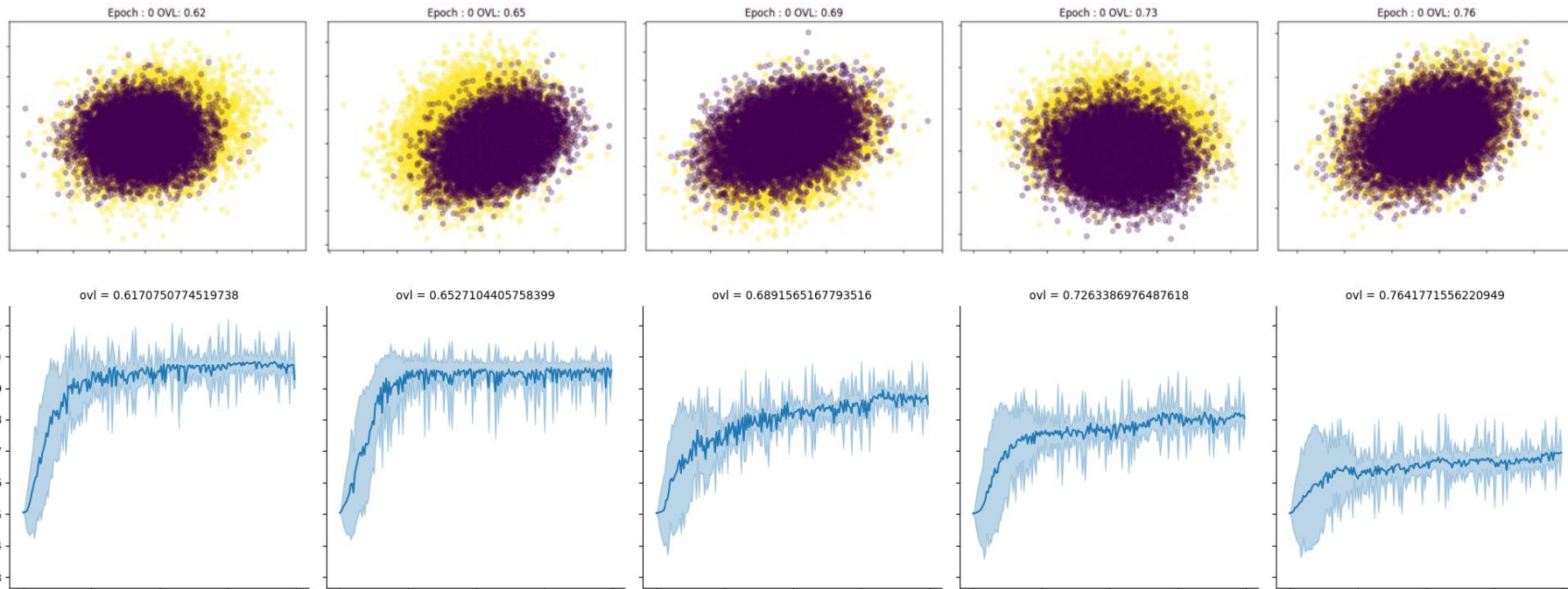
- Variational AutoEncoder
- Loss = reconstruction loss (mean absolute error) + KL divergence
- 200 Epochs
- 10 Runs
- 10 OVL profiles
  - 0.62, 0.65, 0.69, 0.73, 0.76, 0.80, 0.84, 0.88, 0.92, 0.9
- Batch size 1024
- Encoder layers 512, 256, 128, 2
- Decoder layers 2, 128, 256, 512
- Glorot uniform weight initialization
- Adagrad optimizer
- Learning rate 0.01
- Weight regularizer  $10^{-7}$



# Separation Visualisations

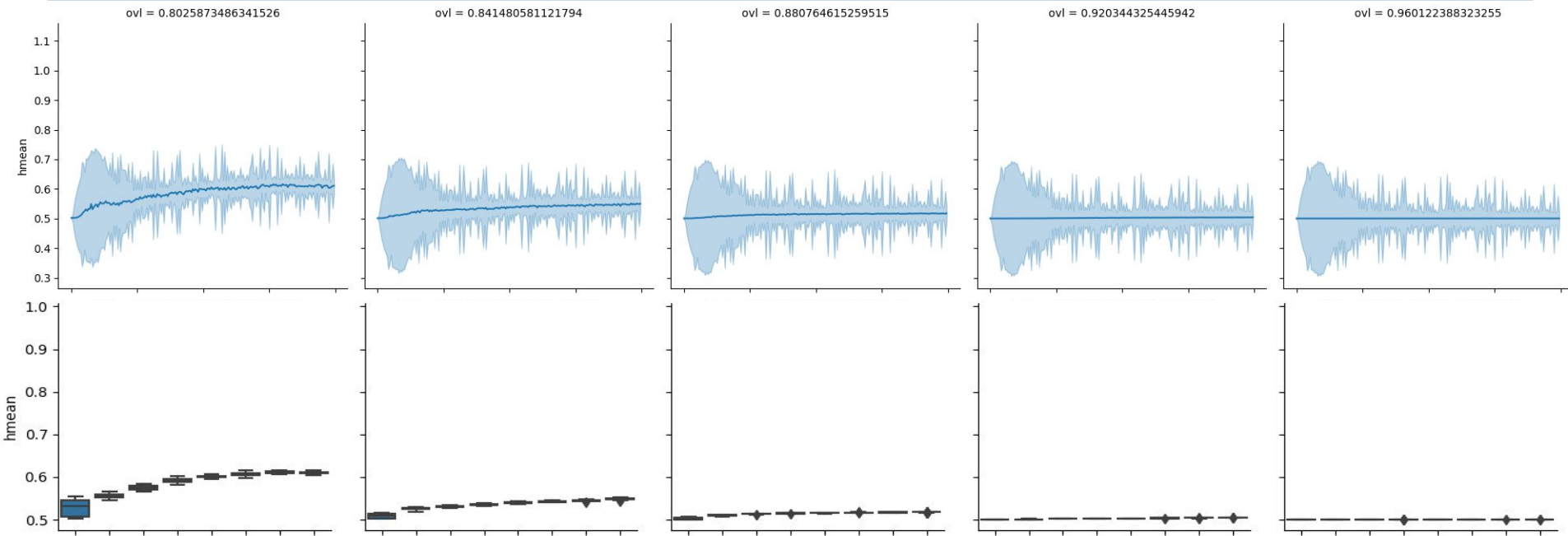
- Separation in the latent space sometimes reaches 1.0 within 50 - 100 iterations for OVL < ~0.65.
- Thereafter, time requirements increase, and separation after 200 epochs fall quickly to ~0.7 at an OVL of 0.76.



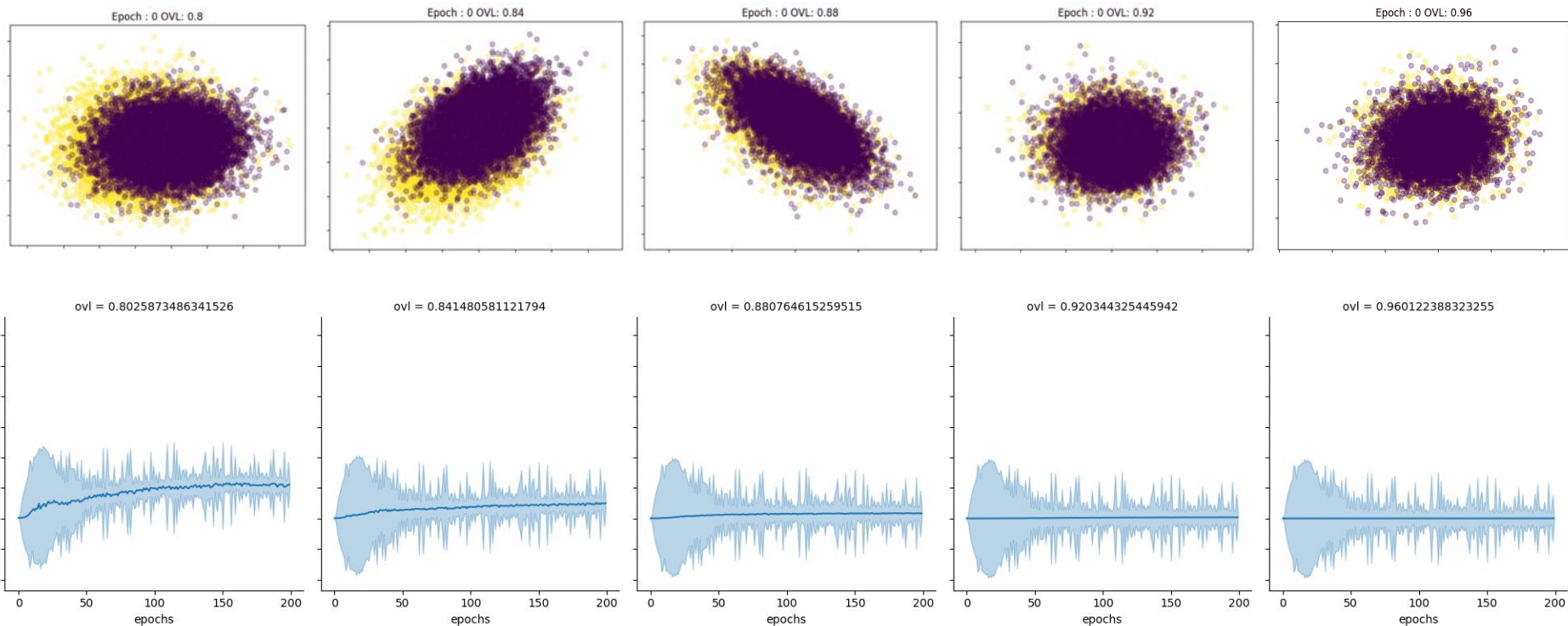


- Assuming two gaussian distribution and sufficient density, a latent separation of 0.7 would allow us to isolate a dense good region in the latent space in an OMNI implementation.

- At OVL > 0.8, latent separation suffers dramatically, rarely reaching values above 0.6.
- Above an OVL of 0.9, the model is unable to separate the data at all.



# Separation vs Epochs



- OVL values above 0.9 often produce latent configurations where the good group is mostly eclipsed by the bad group. (OVL = 0.92 looks very similar to results we have observed).



# Conclusions

- We have defined a sound metric to quantify separability in the latent space.
- As expected, OVL has a major effect on the separability of data groups in the latent space.
- For Gaussian data groups, an OVL of  $< 0.7$  gives adequate separation to isolate the densest region of the good data group
- For higher OVLs, any isolated good regions would likely not be very dense, since the majority of the data is eclipsed by the bad group.

- The effect of the two objective terms (KL divergence and reconstruction loss) is currently being investigated. Results are not yet available.
  - The same experimental setup is used, but additional measurements are taken at each epoch:
    - Overall objective value
    - Value of the KL divergence term
    - Value of the reconstruction term
- We do not yet understand the interaction between the actual distribution of the input data and the prior distribution enforced on the latent space.
  - Experiments to investigate this interaction still need to be formalised.
- Future experiments may include:
  - Investigation of information conservation/encoding in individual layers of the encoder network.
- The experiments code can be made available to the company to provide a flexible framework for model implementation