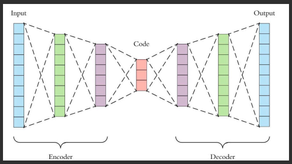
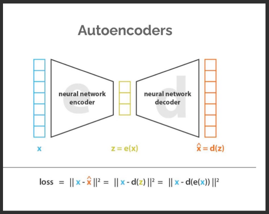
AUTOENCODERS - AE

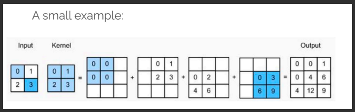
- self-supervised algorithm (labels generated from input)
- reconstruct input after compressing it.



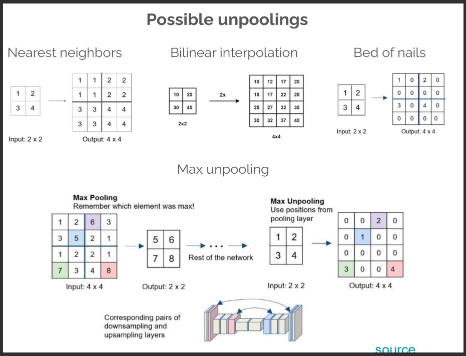


Transposed convolution

- increase size of an image or feature map

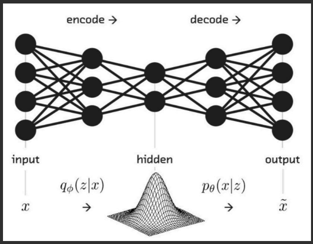


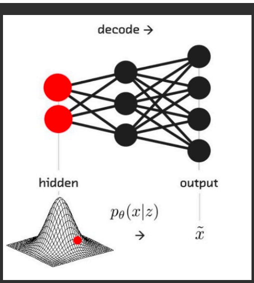
Unpooling

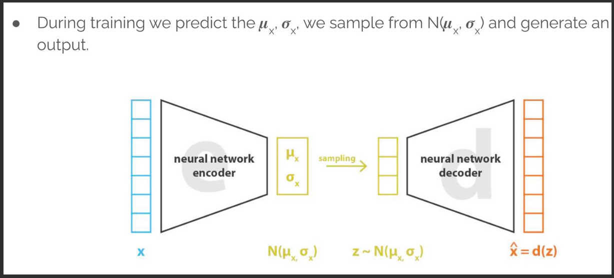


Variational autoencoder - VAE

- Like AE, but you learn parameters of a distribution modeling your data.
- In the middle of the network, the latent space, similar points encode a similar inputs
- at inference time you sample points from the distribution and pass to the decoder, it generates a new image not in the TR set.
- generative models







The **total loss** is a combination of the two:

$$J_{\text{VAE}} = \boxed{J_{\text{reconstruction}}} + \beta \boxed{J_{\text{KL}}}$$

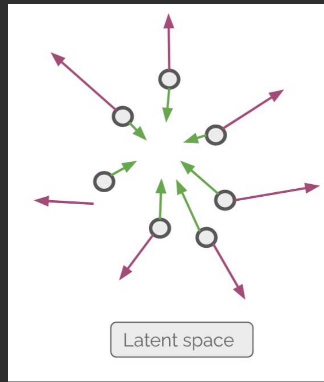
where β is a weighting factor that controls the trade-off between reconstruction and regularization.

The two losses **compete** with each other:

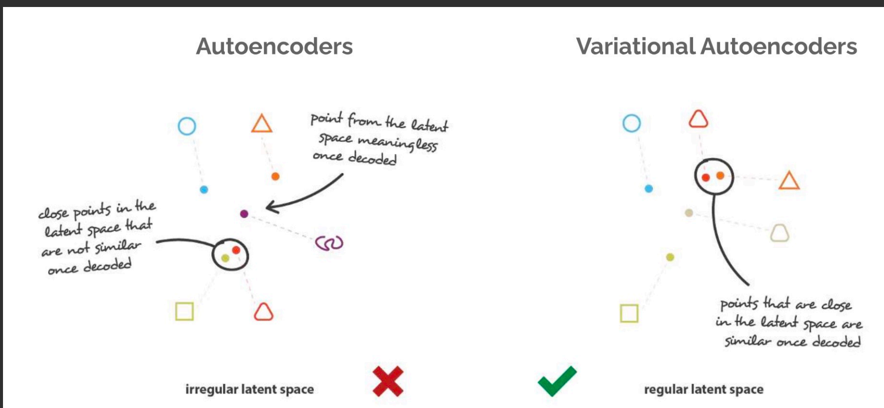
- **Reconstruction Loss wants to separate points** to improve decoding accuracy.
- **KL Loss wants to group points together** to ensure a continuous, smooth latent space.

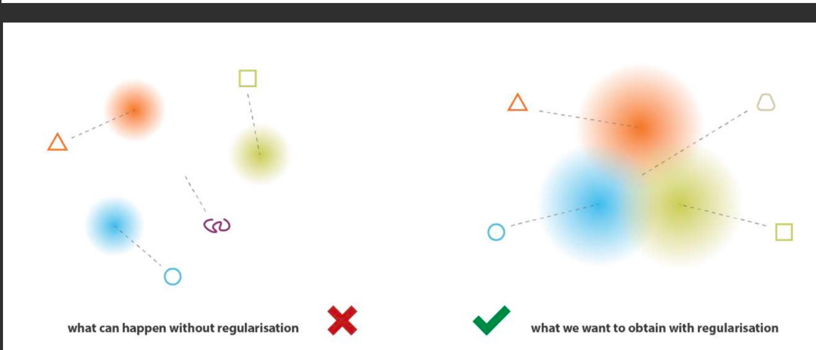
If β is too **low** → The model ignores regularization and memorizes the training data (like a standard Autoencoder).
If β is too **high** → The model learns a smooth latent space but with poor reconstruction quality.

$$\text{loss} = \| x - \hat{x} \|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = \| x - d(z) \|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)]$$



Latent space differences





🔴 **What happens?**

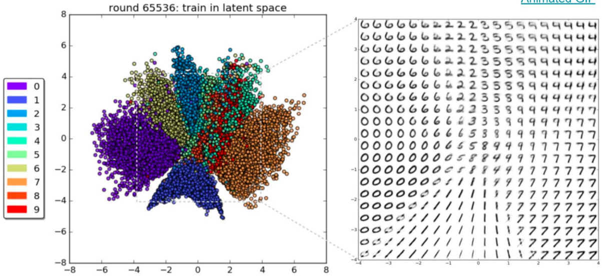
- The latent representations (colored blobs) are **scattered** and do not overlap.
- Each point is encoded into a **separate region** of the latent space.
- The space is **sparse**, making interpolation between points difficult.
- This can lead to **overfitting** and poor generalization when sampling new data.

🟢 **What happens?**

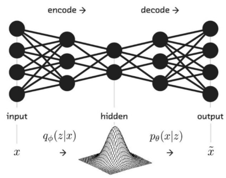
- The latent space **clusters better** with smooth transitions between different groups.
- Different representations **overlap**, forming a **continuous** distribution.
- The model **generalizes better**, meaning it can generate smooth interpolations and meaningful new samples.

Variational Autoencoders - Latent space reconstruction

[Animated GIF]



- Variational Autoencoders (VAE) are good at learning representations (interpretable dimensions, possibility to set complex priors).



- Generative Adversarial Networks (GANs) are good at generating new samples (clever loss). Trickier to train.