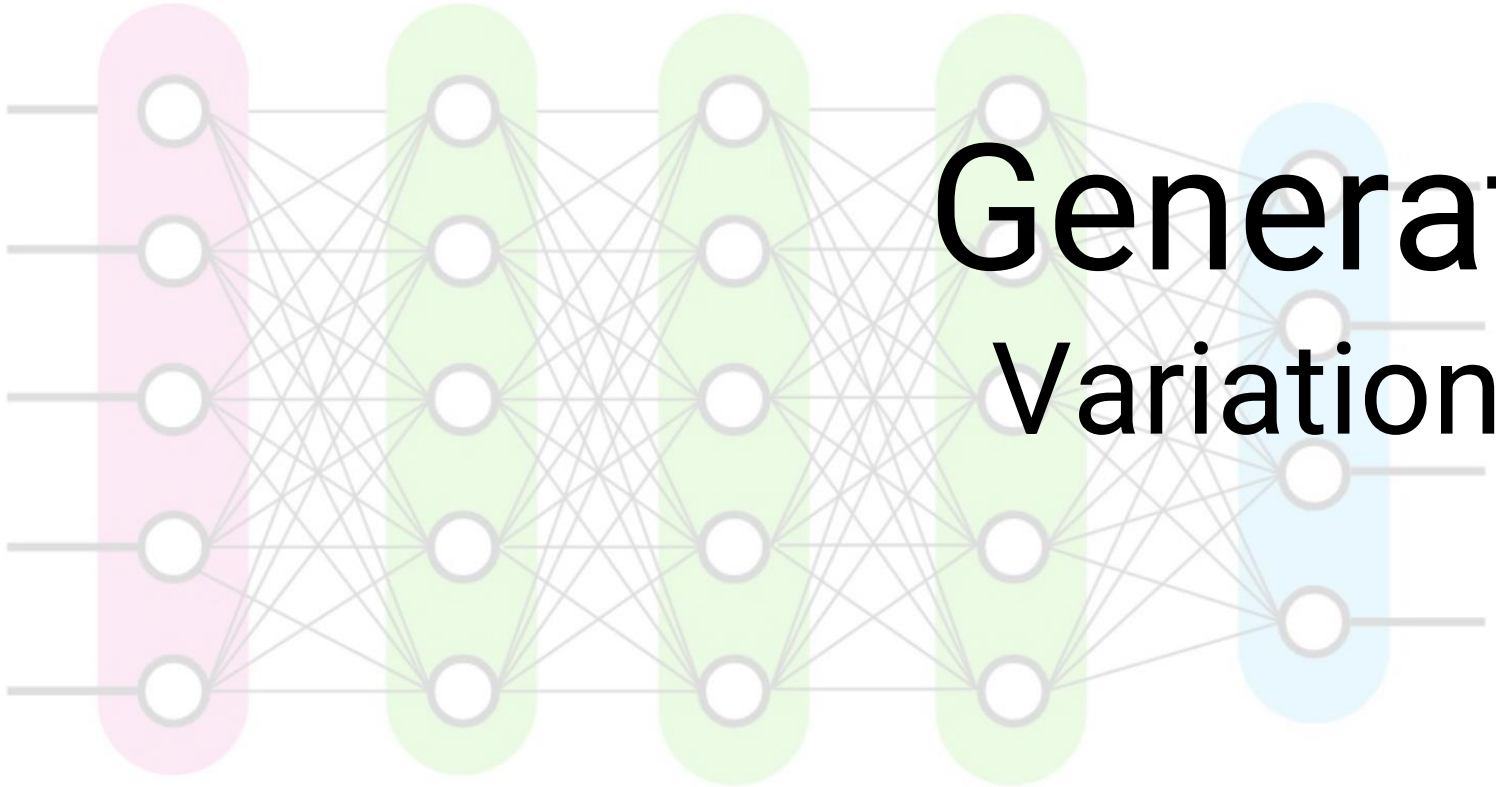
Class core values

1. Be **respect**ful to yourself and others
2. Be **confident** and believe in yourself
3. Always do your **best**
4. Be **cooperative**
5. Be **creative**
6. Have **fun**
7. Be **patient** with yourself while you learn
8. Don’t be shy to **ask “stupid” questions**
9. Be **inclusive** and **accepting**

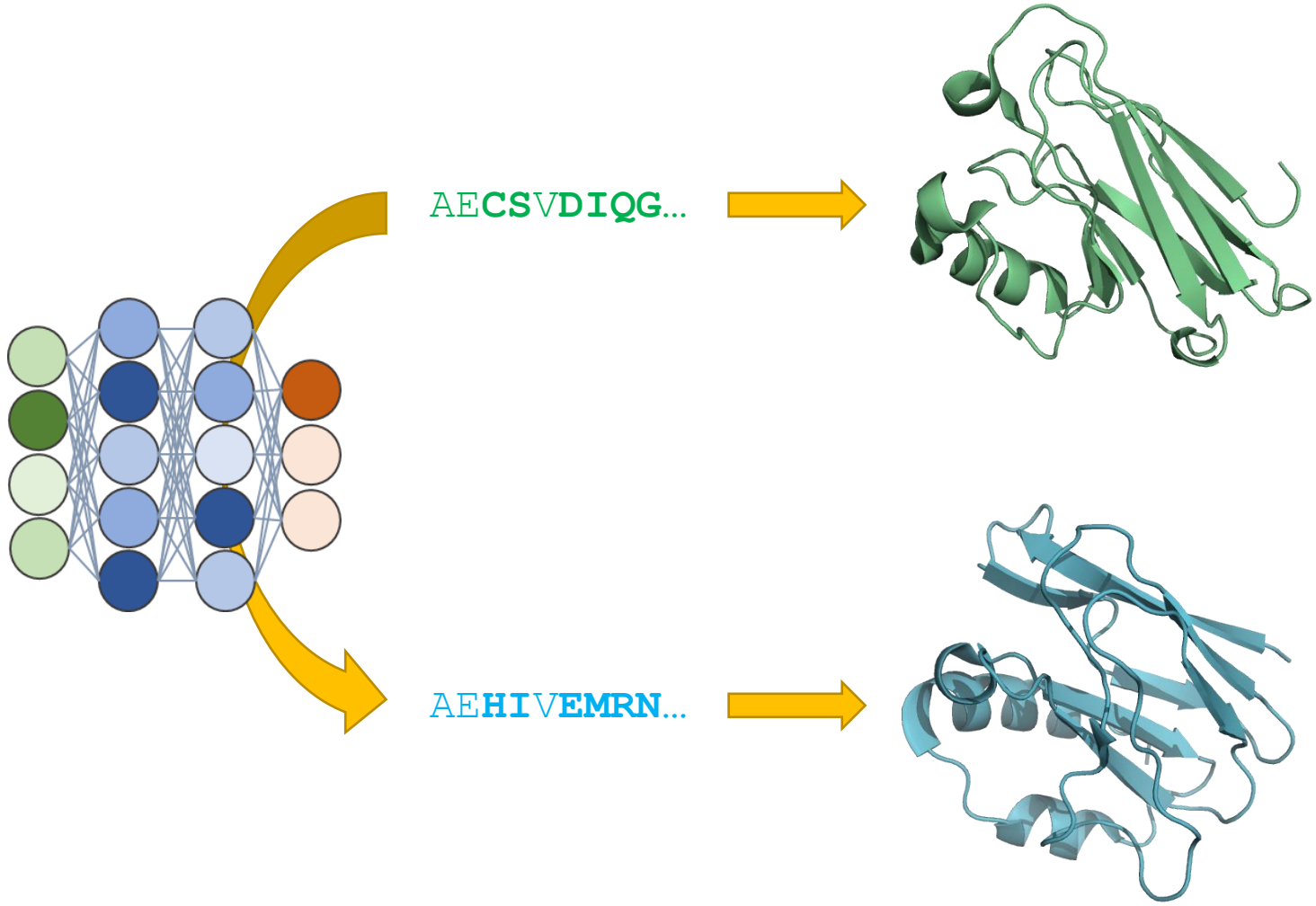


Week 8, Lecture 1

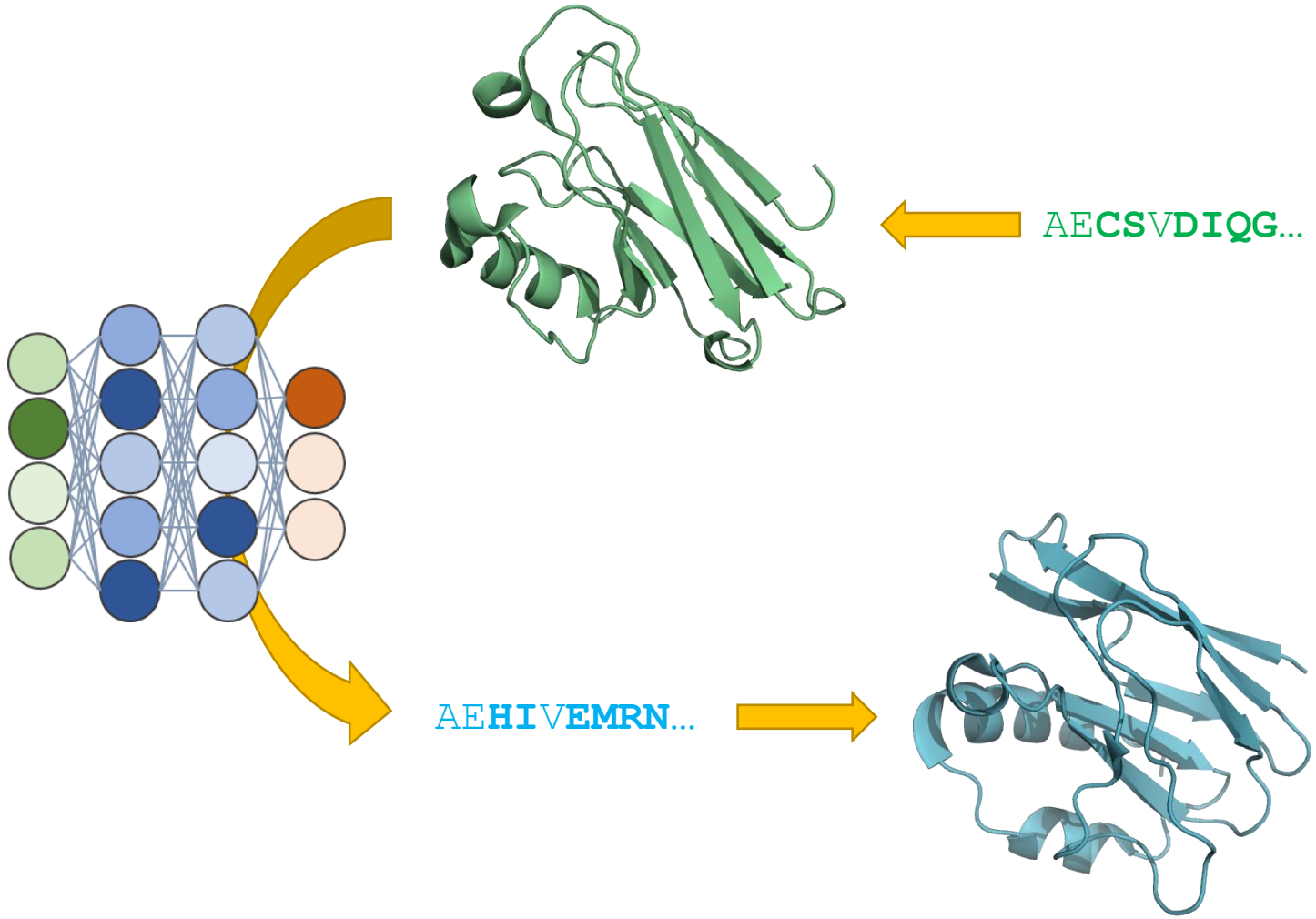
Generative Models:

Variational Autoencoders

Generating proteins

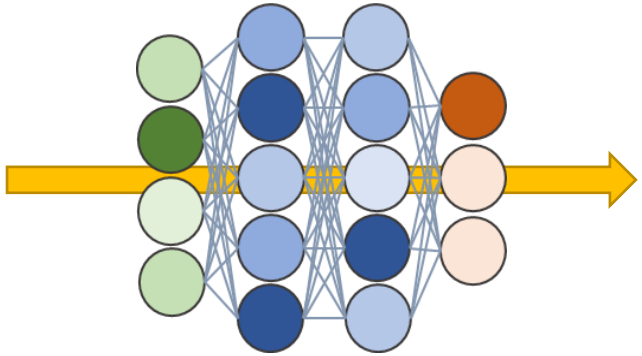


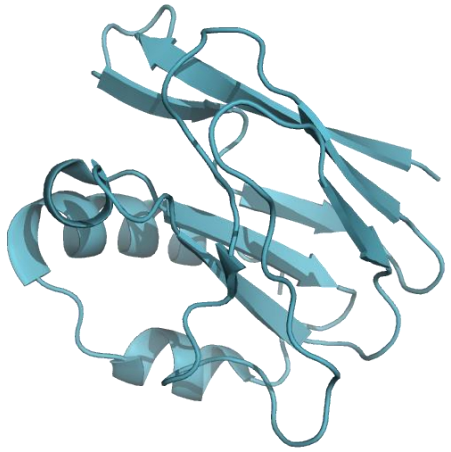
Generating proteins



Uses for a generative protein model?

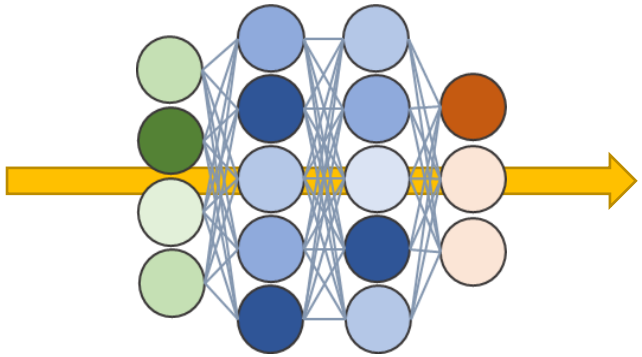
AE**CS**V**DIQG**…





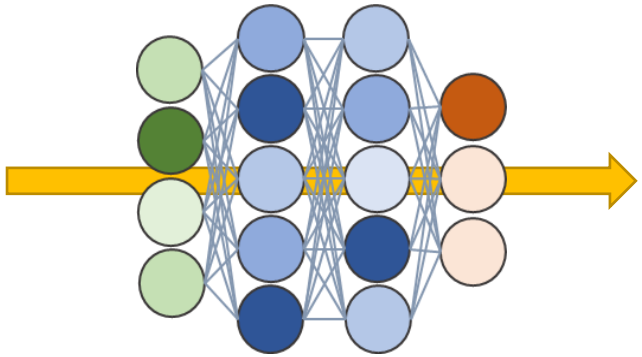
AE**HI**V**EMRN**…

Generative models



Generative models

The Mona Lisa is a half-length portrait painting by Italian artist Leonardo da Vinci.



Girl with a Pearl Earring is an oil painting by Dutch Golden Age painter Johannes Vermeer.

Variational autoencoder (VAE)

Image from [vae-celebA]

Autoencoder

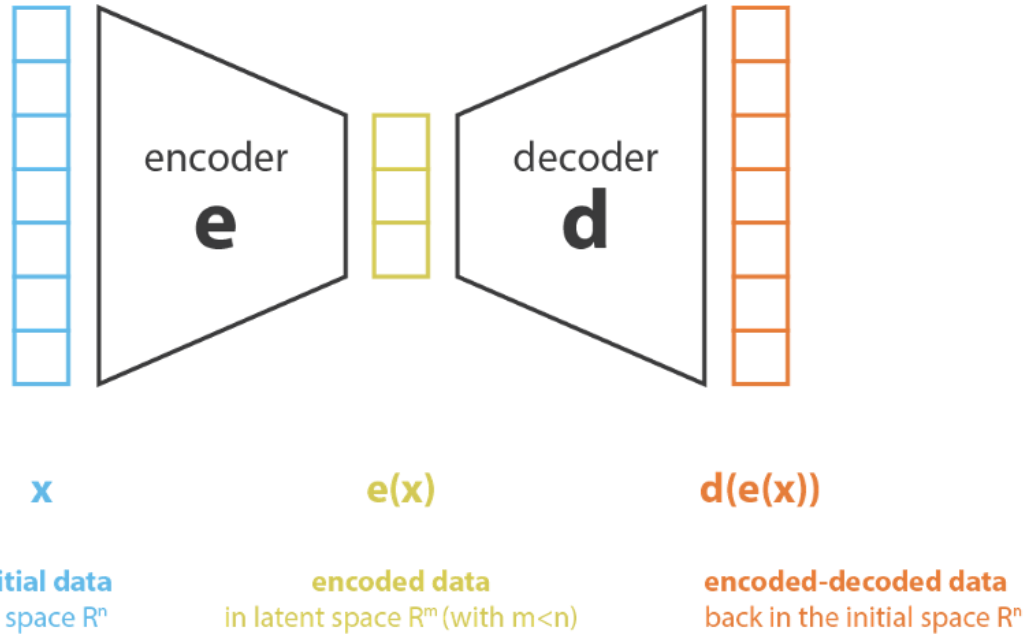
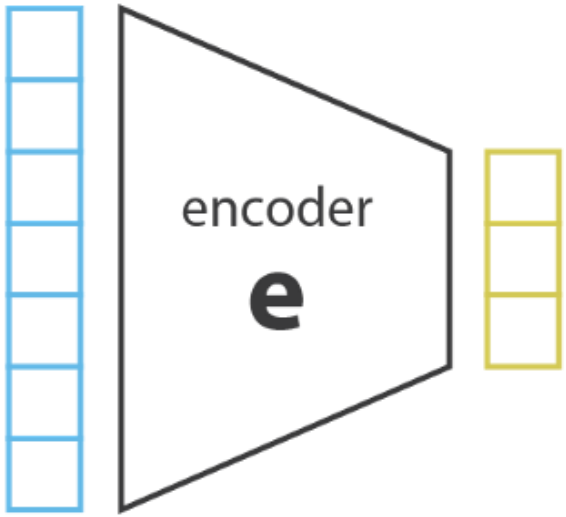


Image from [TDS – Understanding VAEs]

Encoder



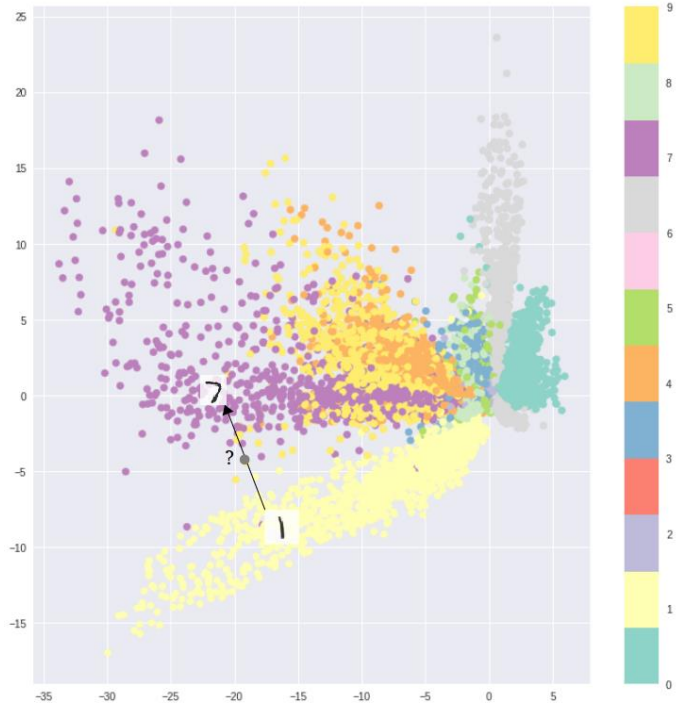
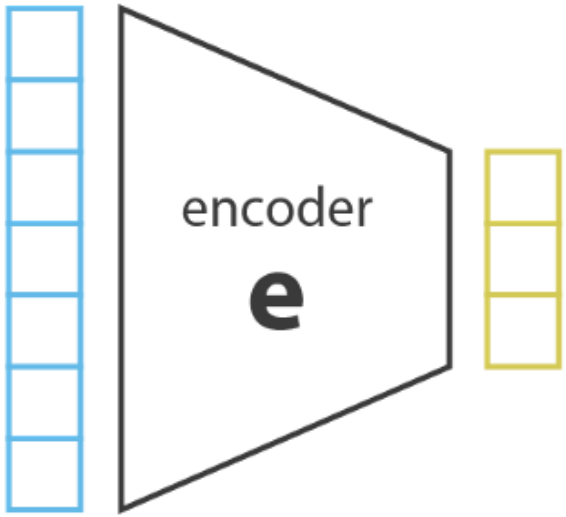
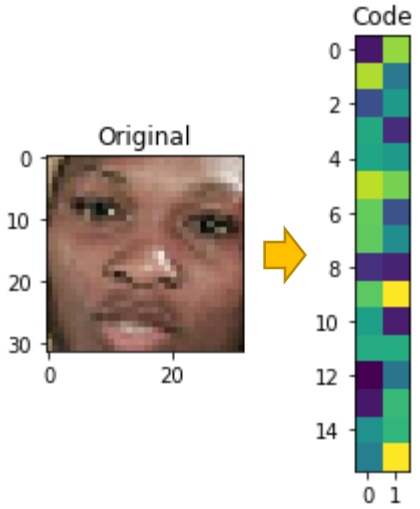


Image from [TDS – Intuitively understanding VAEs]

Encoder





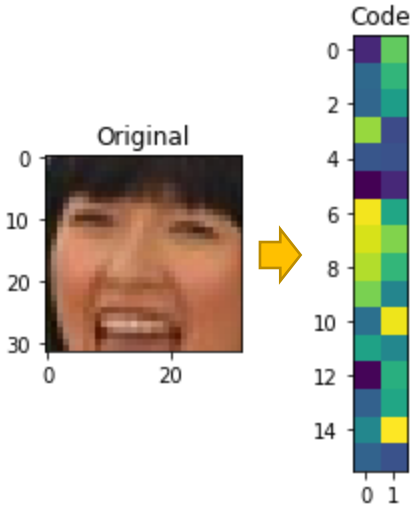
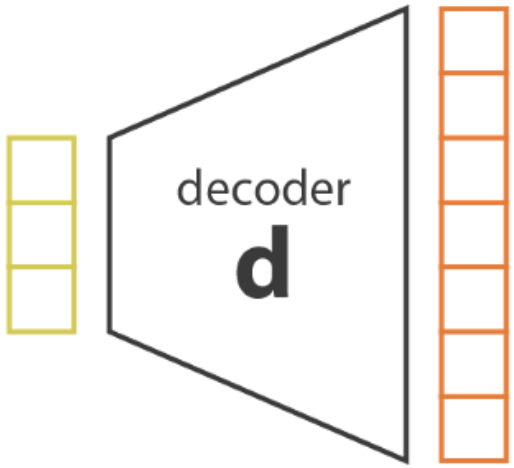


Image from [Autoencoders for Image Reconstruction in Python and Keras]

Decoder



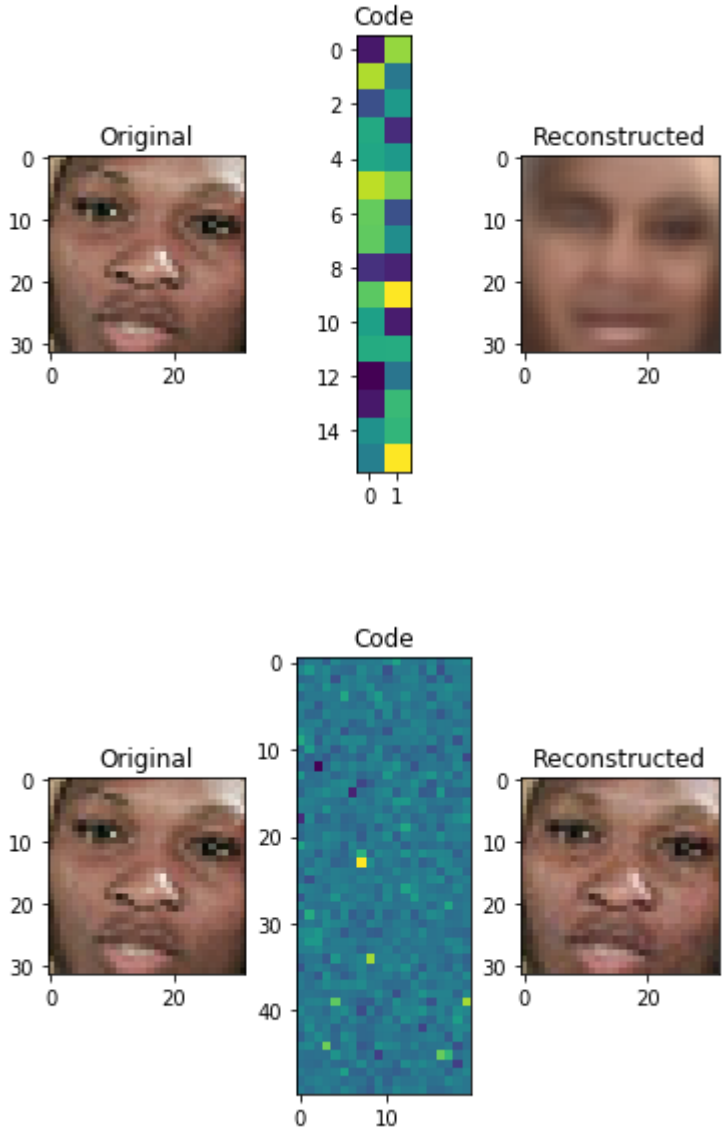
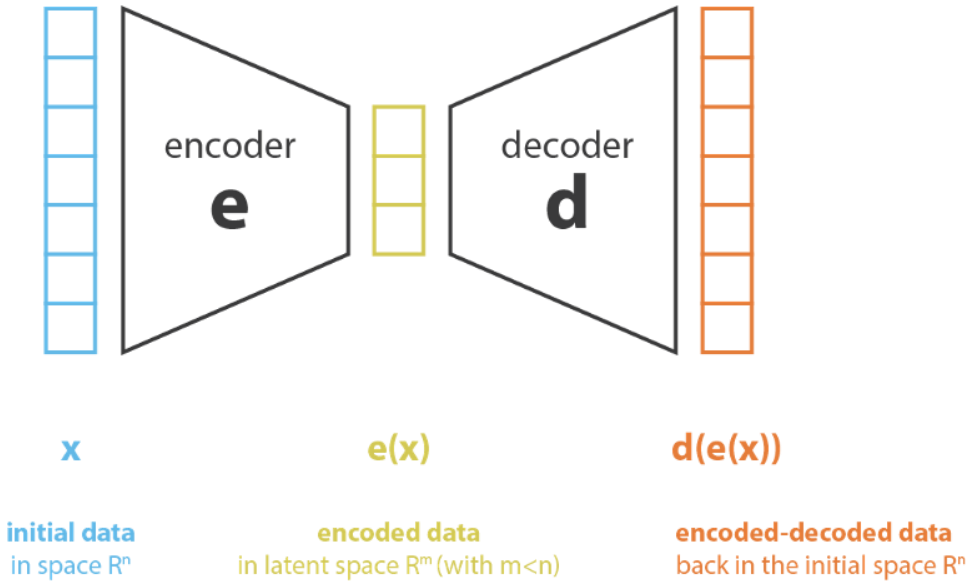
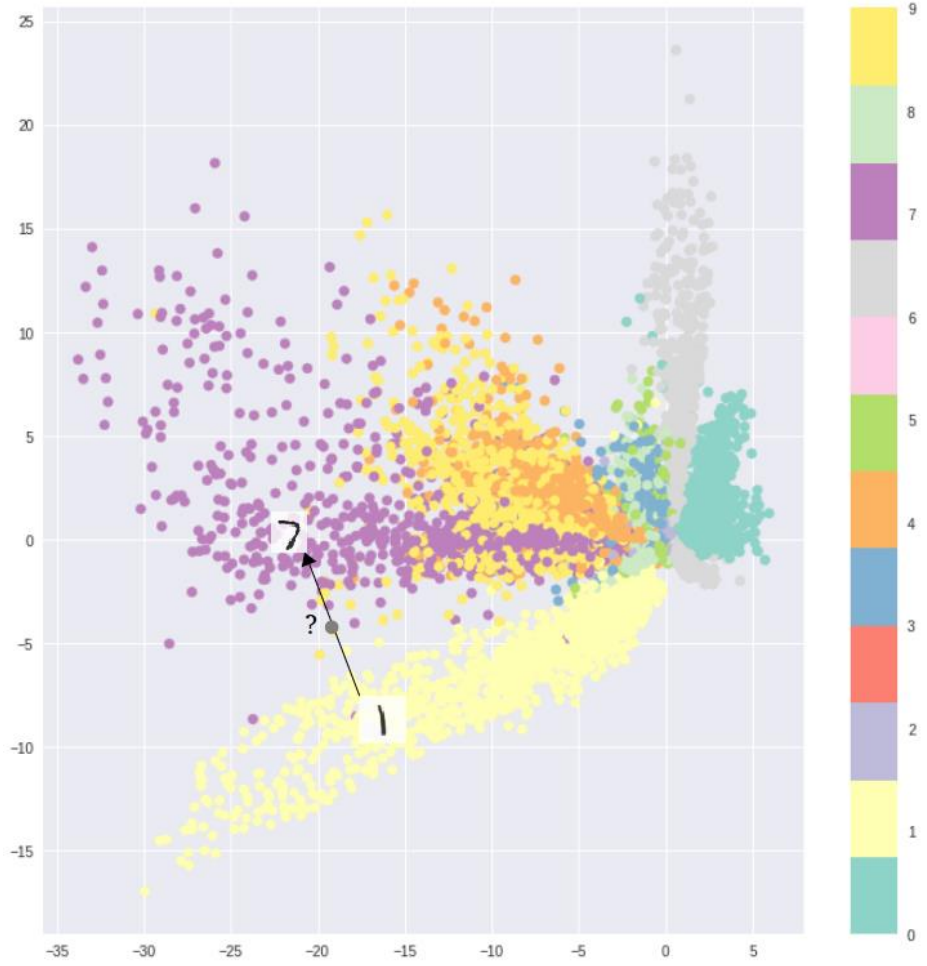


Image from [Autoencoders for Image Reconstruction in Python and Keras]

Latent Space

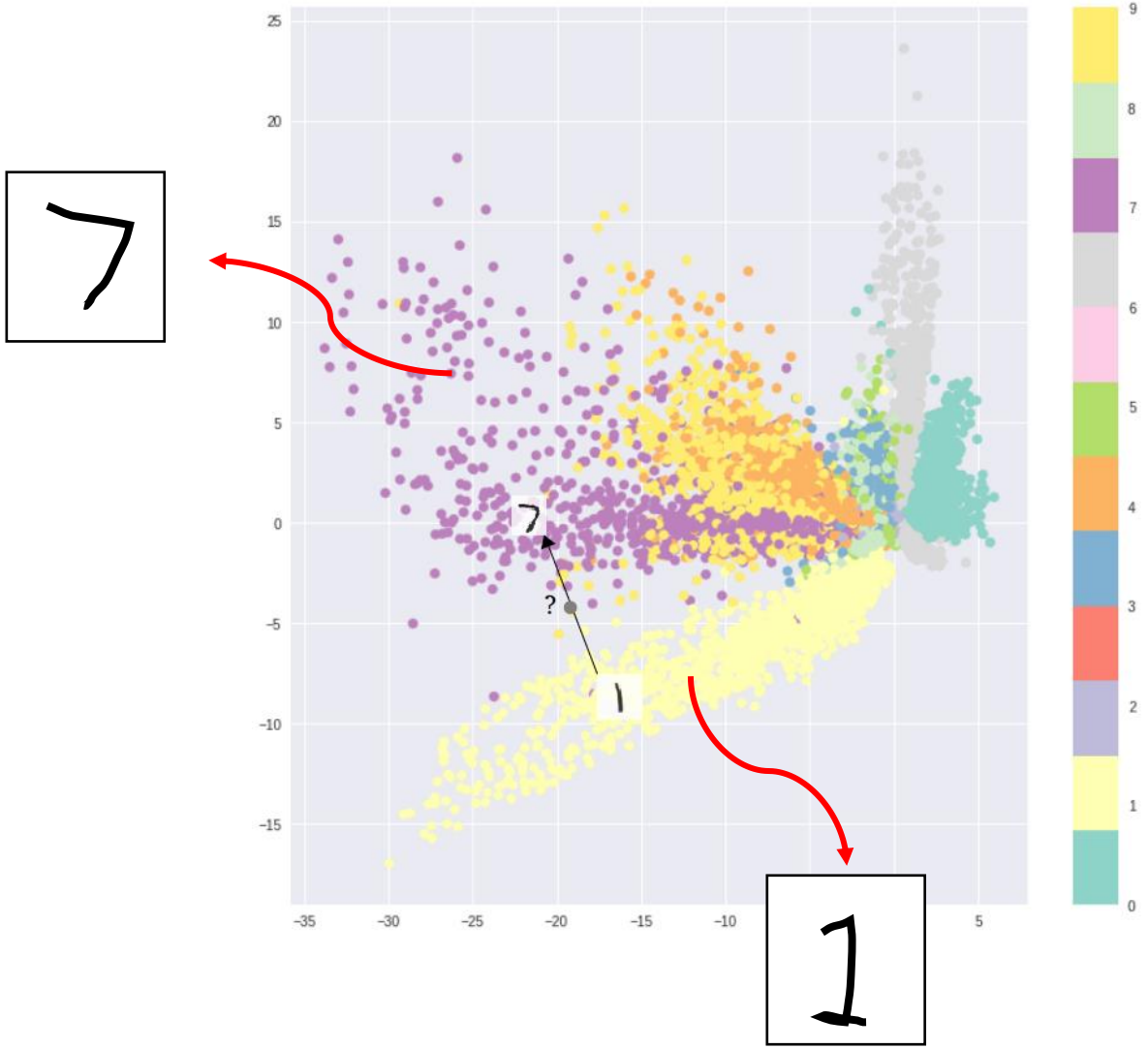




What would happen if we input embedded values from the latent space into the decoder?

Image from [TDS – Intuitively understanding VAEs]

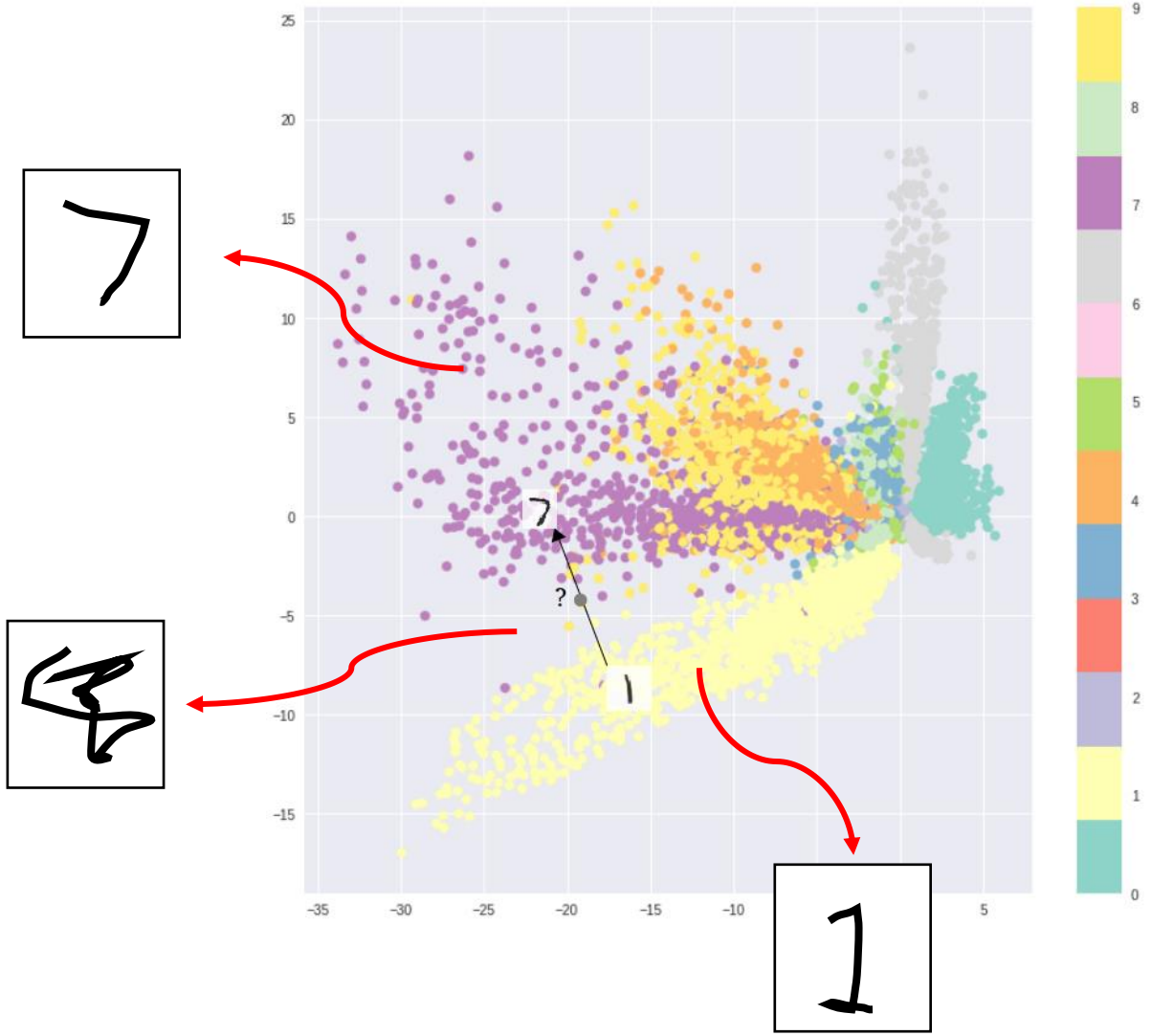
Irregular latent space of autoencoders



What would happen if we input random values from the latent space into the decoder?

Image from [TDS – Intuitively understanding VAEs]

Irregular latent space of autoencoders

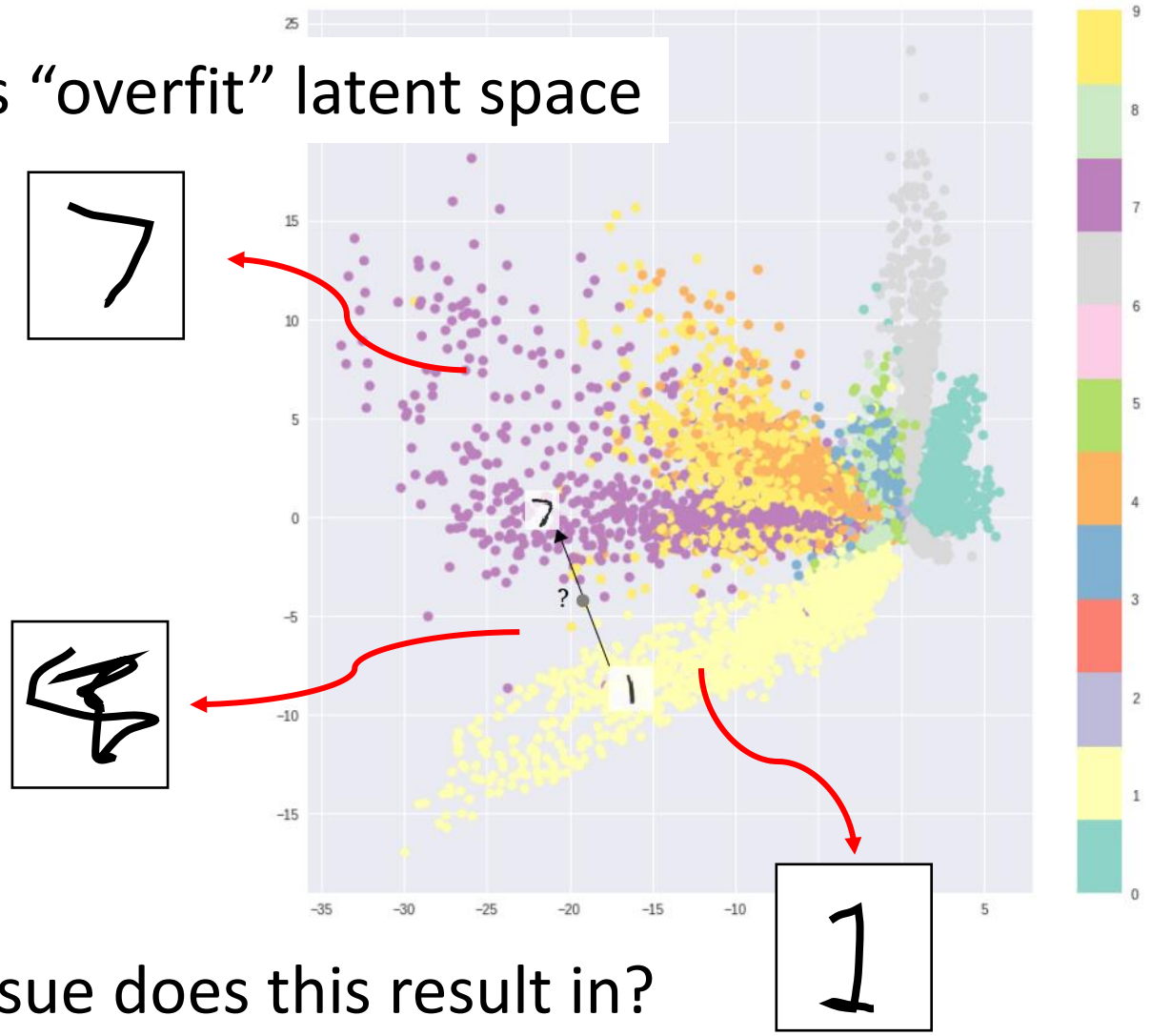


Random values from the latent space decode to meaningless data

Image from [TDS – Intuitively understanding VAEs]

Irregular latent space of autoencoders

Autoencoders “overfit” latent space

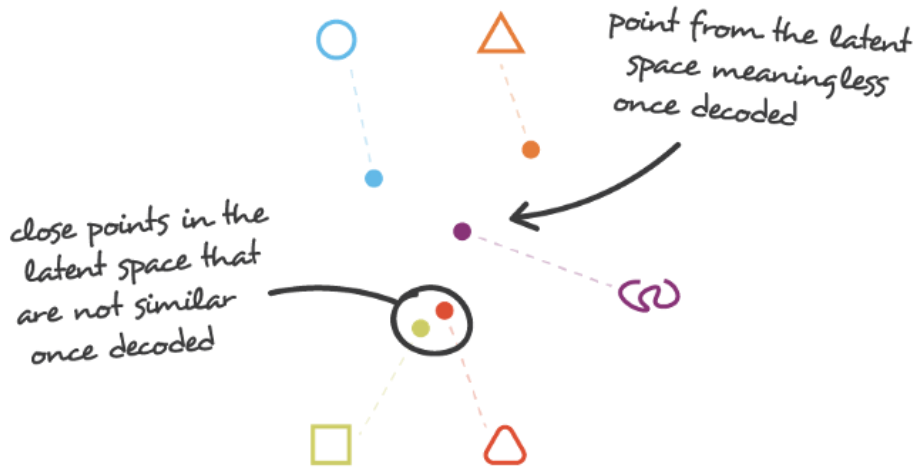


Random values from the latent space decode to meaningless data

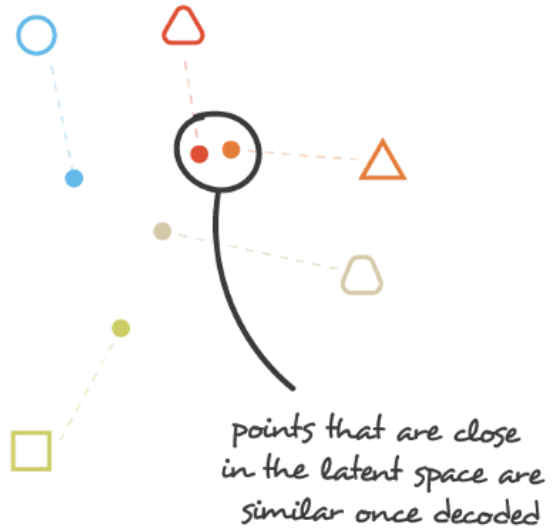
What other issue does this result in?

Image from [TDS – Intuitively understanding VAEs]

Irregular latent space of autoencoders



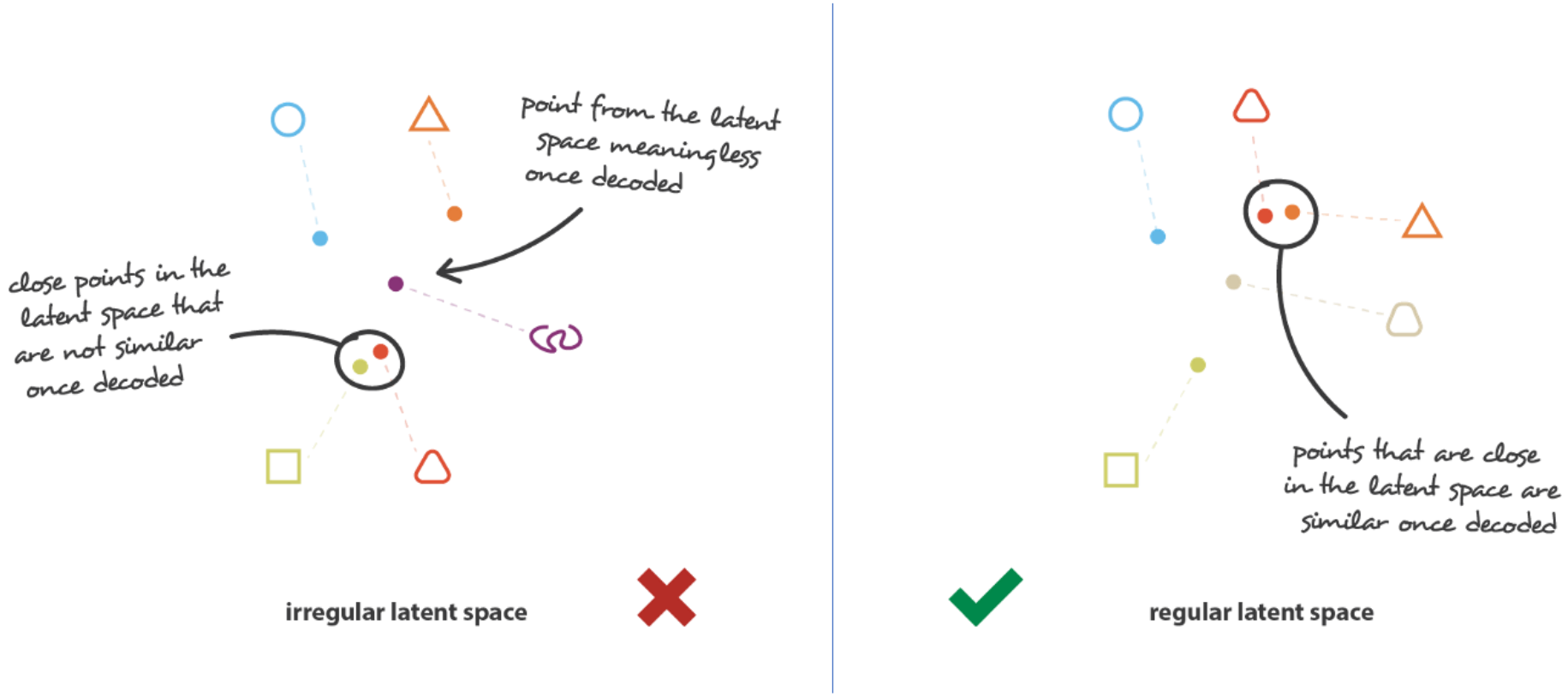
irregular latent space



regular latent space

Image from [TDS – Understanding VAEs]

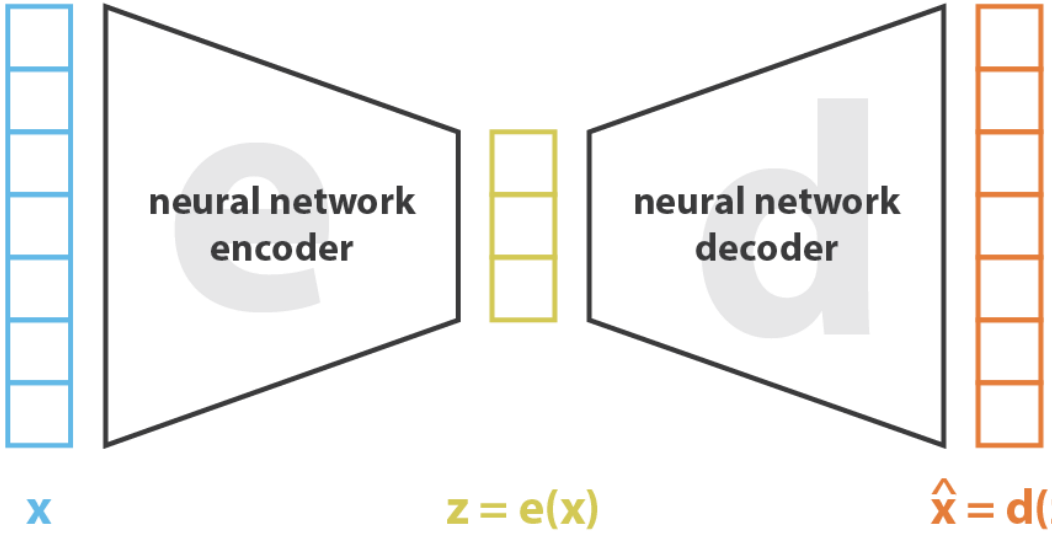
Autoencoders cannot be used to generate new data



What needs to be changed to generate new data?

Image from [TDS – Understanding VAEs]

Autoencoder loss function doesn't care about latent space

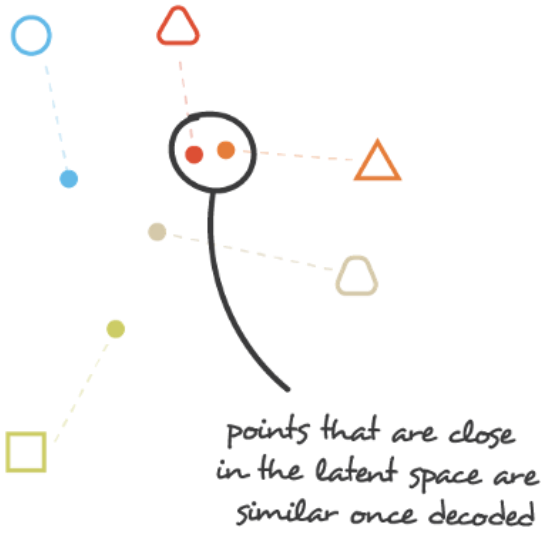


$$\text{loss} = \| x - \hat{x} \|^2$$

Need to make some adjustments to the loss function

Image from [TDS – Understanding VAEs]

Data is encoded as a distribution instead of single point



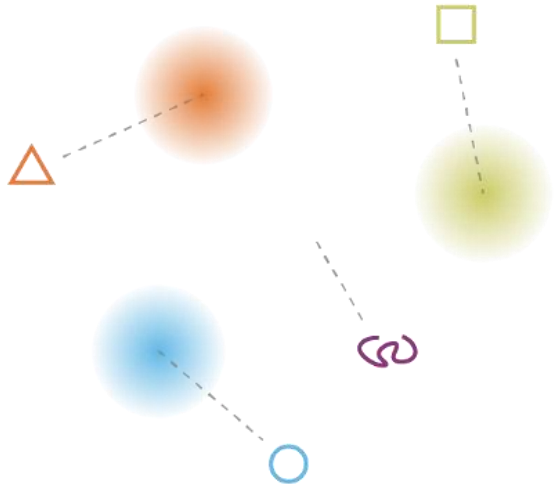
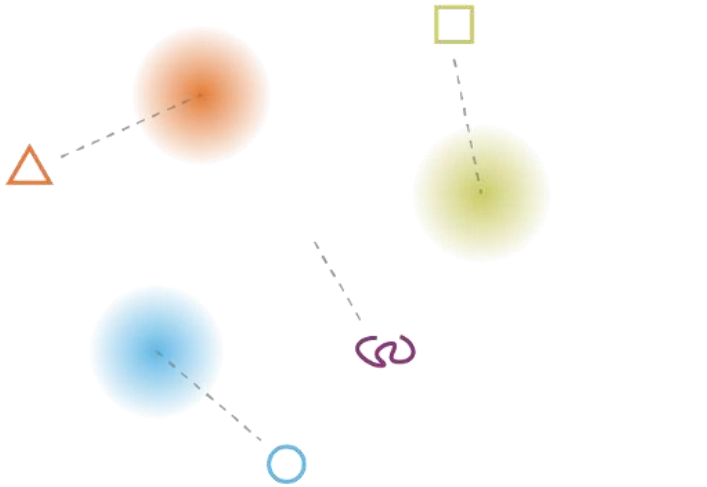
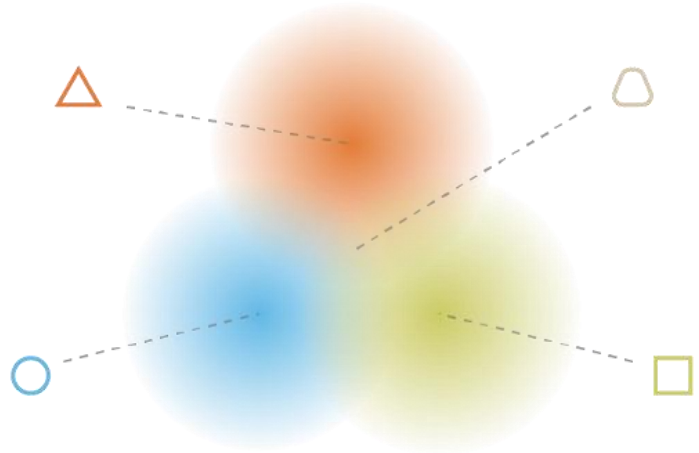


Image from [TDS – Understanding VAEs]

Regularization for latent space (mean and variance)



what can happen without regularisation



what we want to obtain with regularisation

Image from [TDS – Understanding VAEs]

Regularization for latent space (mean and variance)

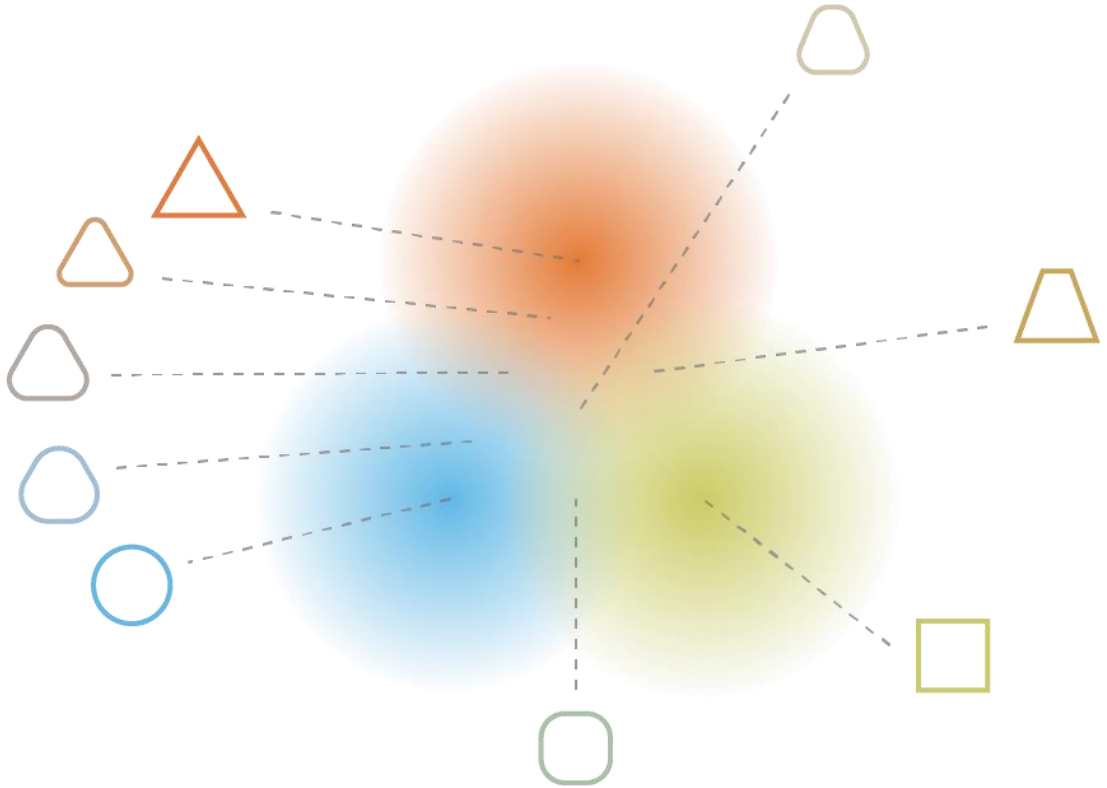
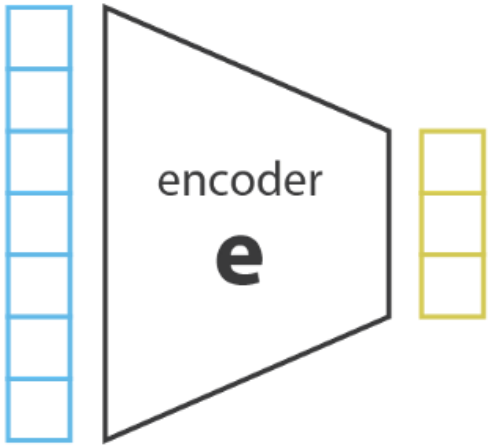
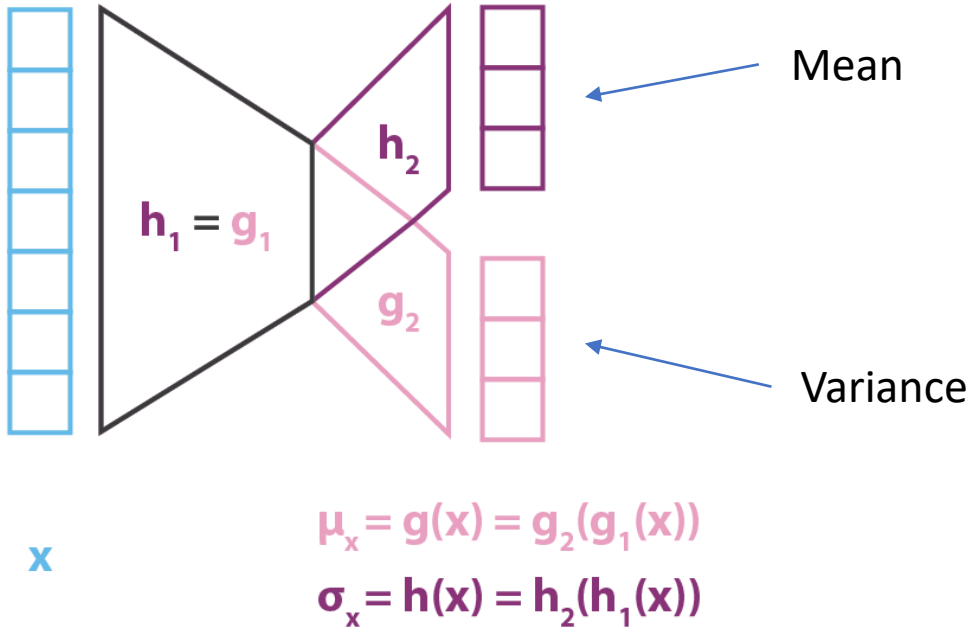


Image from TDS – Understanding VAEs

Regularization for latent space (mean and variance)

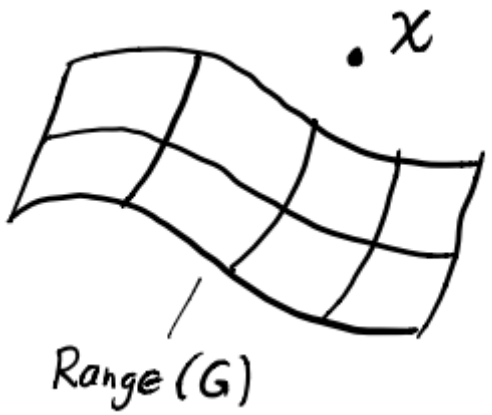




Loss function and backpropagation?

Image from [TDS – Understanding VAEs]

Variational inference



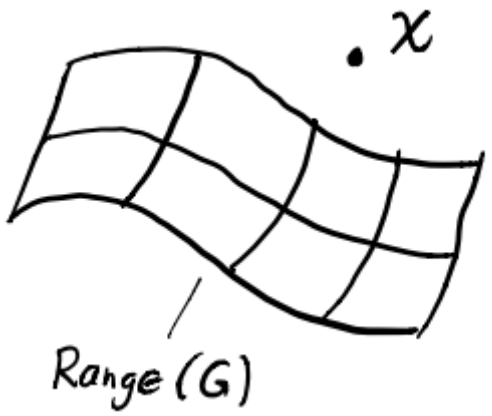
$$p(z|x) = \frac{p(x|z)\,p(z)}{p(x)}$$

$$p(x) = \int p(x|z)\,p(z)\,dz$$

Problem: intractable

Image from NEU CS7150: VAEs

Variational inference



$$p(z|x) = \frac{p(x|z)\,p(z)}{p(x)}$$

$$p(x) = \int p(x|z)\,p(z)\,dz$$

Problem: intractable

Estimate

$$q_x(z) \equiv \mathcal{N}(g(x), h(x))$$

Solution: tractable

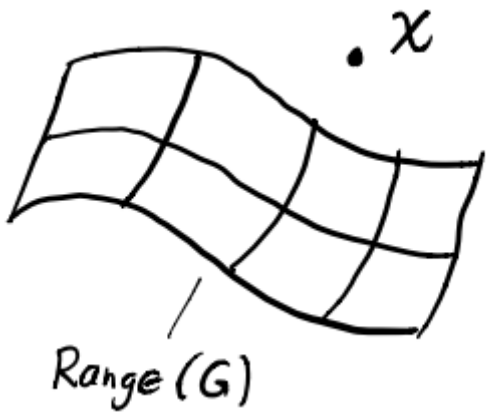
Functions (e.g., Gaussians)

$g \in G$

$h \in H$

Image from NEU CS7150: VAEs

Variational inference



$$p(z|x) = \frac{p(x|z)\,p(z)}{p(x)}$$

$$p(x) = \int p(x|z)\,p(z)\,dz$$

Problem: intractable

Estimate

Functions (e.g., Gaussians)

$$q_x(z) \equiv \mathcal{N}(g(x), h(x)) \qquad g \in G \qquad h \in H$$

Solution: tractable

Optimizing lower bound of probability distribution

$$\begin{aligned}
(g^*, h^*) &= \underset{(g,h)\in G\times H}{\arg\min}\; KL(q_x(z), p(z|x)) \\
&= \underset{(g,h)\in G\times H}{\arg\min} \left(\mathbb{E}_{z\sim q_x}(\log q_x(z)) - \mathbb{E}_{z\sim q_x}\left(\log \frac{p(x|z)p(z)}{p(x)}\right)\right) \\
&= \underset{(g,h)\in G\times H}{\arg\min} \left(\mathbb{E}_{z\sim q_x}(\log q_x(z)) - \mathbb{E}_{z\sim q_x}(\log p(z)) - \mathbb{E}_{z\sim q_x}(\log p(x|z)) + \mathbb{E}_{z\sim q_x}(\log p(x))\right) \\
&= \underset{(g,h)\in G\times H}{\arg\max} \left(\mathbb{E}_{z\sim q_x}(\log p(x|z)) - KL(q_x(z), p(z))\right) \\
&= \underset{(g,h)\in G\times H}{\arg\max} \left(\mathbb{E}_{z\sim q_x}\left(-\frac{||x - f(z)||^2}{2c}\right) - KL(q_x(z), p(z))\right)
\end{aligned}$$

Image from [NEU CS7150: VAEs]

VAE loss function: reconstruction error and regularization term (KL divergence)

$$E_{q(z|x)} \log p(x|z) - KL(q(z|x) || p(z))$$

Reconstruction error

Regularization term (KL divergence)

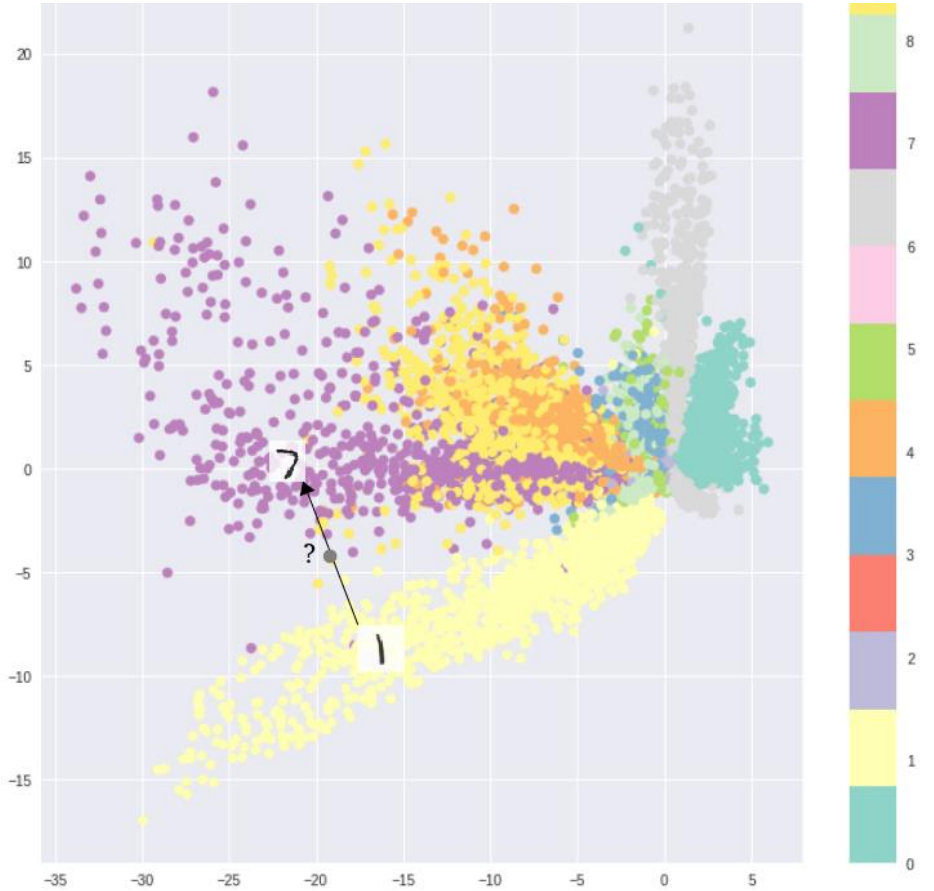
Autoencoder loss = $|| x - \hat{x} ||^2$

Distance between probability distributions

What does this mean for the latent space?

Reconstruction loss only

$$E_{q(z|x)} \log p(x|z) - KL(q(z|x)||p(z))$$

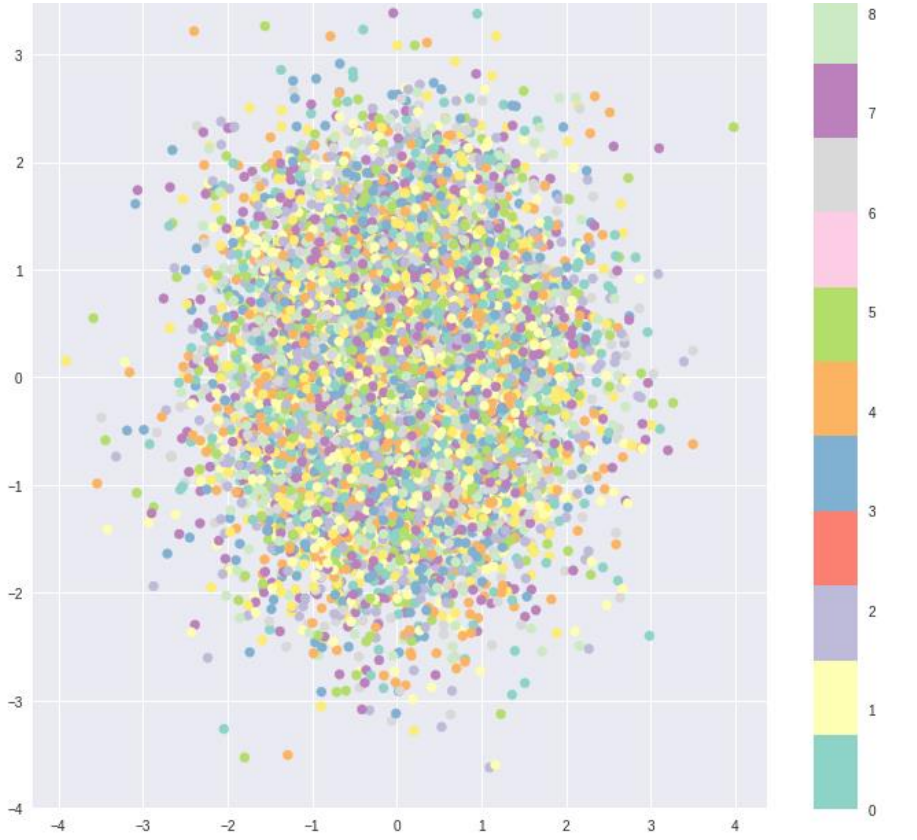


Why does it have this shape?
What about KL only?

Image from [TDS – Intuitively understanding VAEs]

KL Divergence loss only

$$E_{q(z|x)} \log p(x|z) - KL(q(z|x) \,||\, p(z))$$

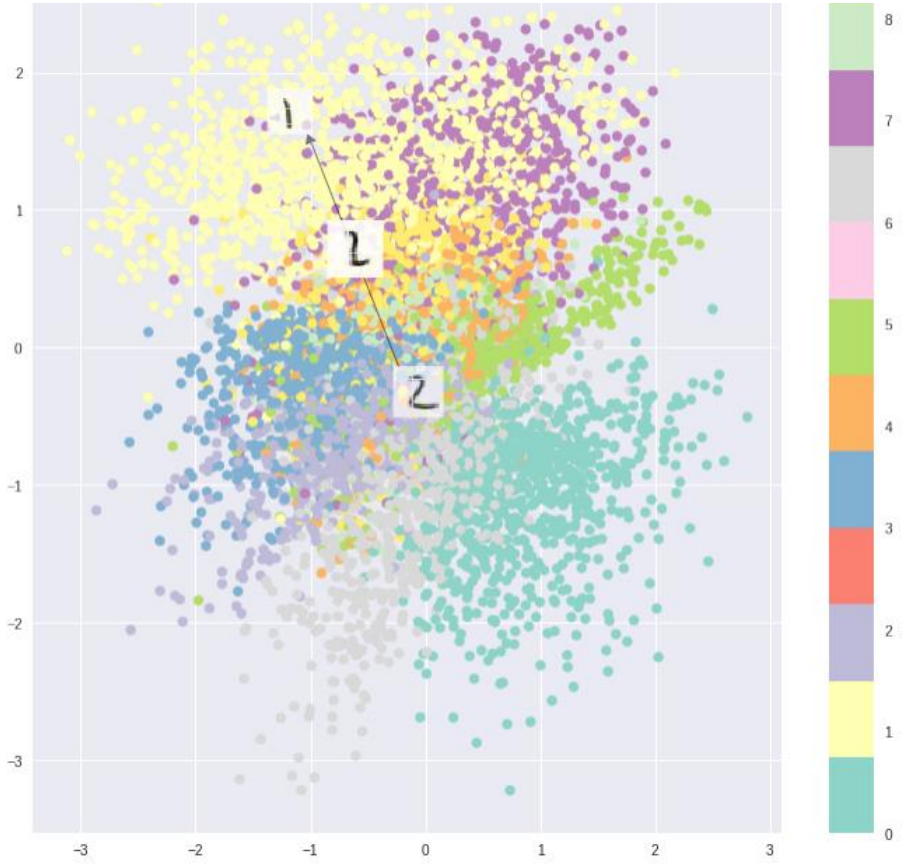


Why does it have this shape?
Reconstruction + KL?

Image from [TDS – Intuitively understanding VAEs]

Reconstruction and KL divergence

$$E_{q(z|x)} \log p(x|z) - KL(q(z|x) || p(z))$$



Why does it have this shape?

Image from [TDS – Intuitively understanding VAEs]

VAE model steps

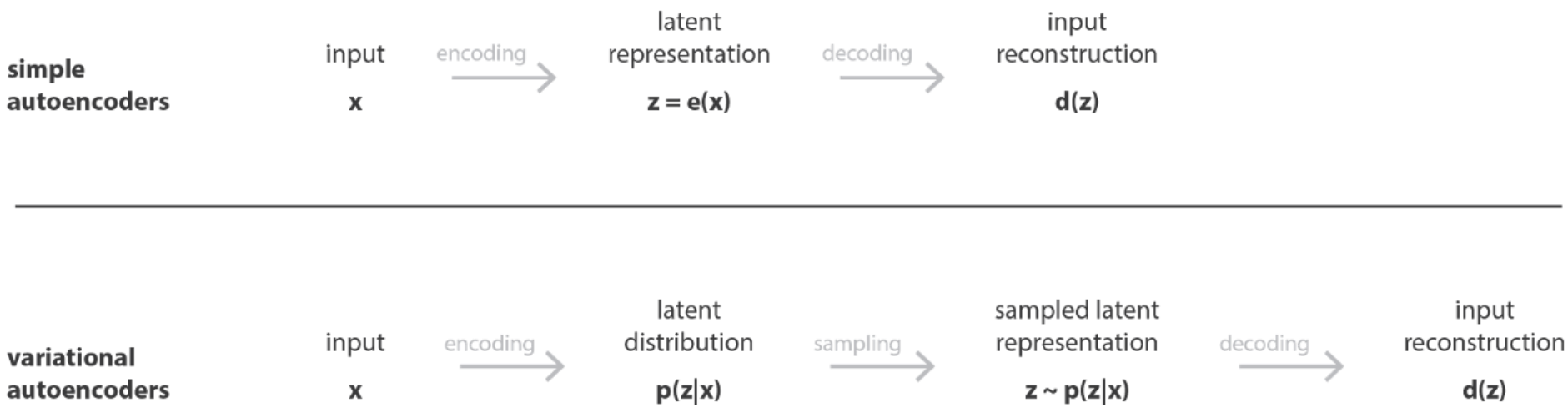
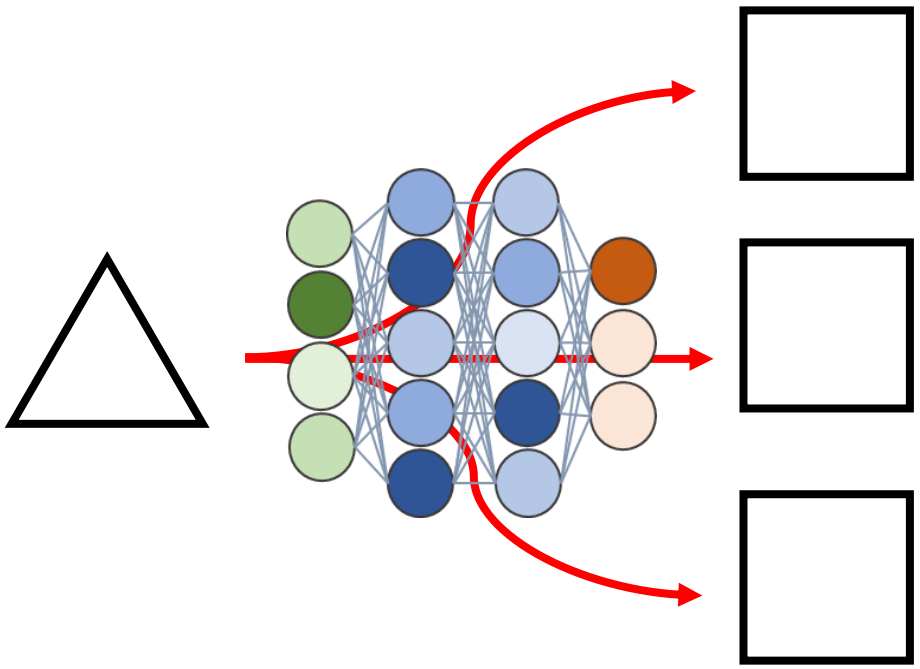


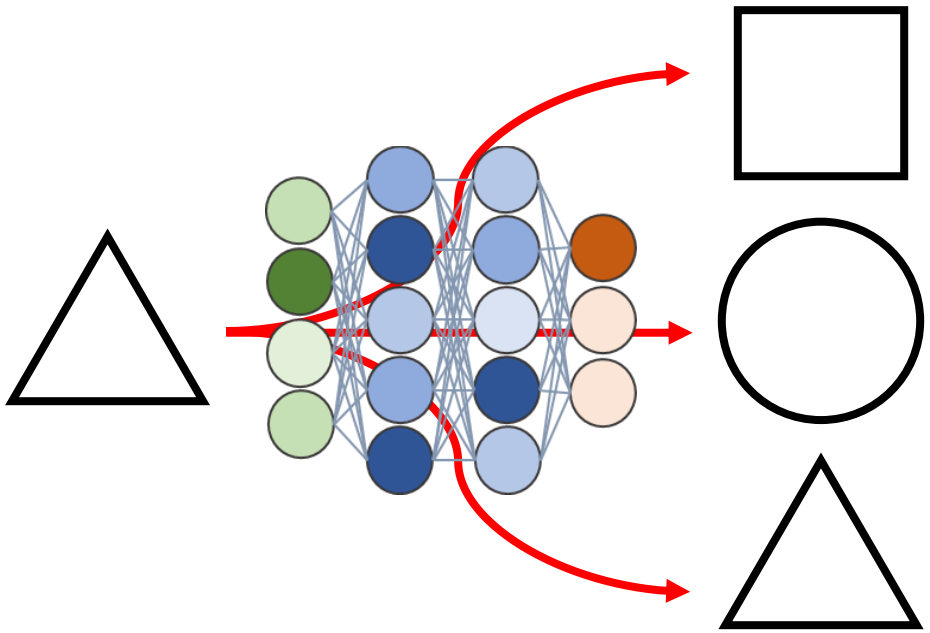
Image from [TDS – Understanding VAEs]

Deterministic vs probabilistic

Deterministic



Probabilistic



Problem: backpropagation

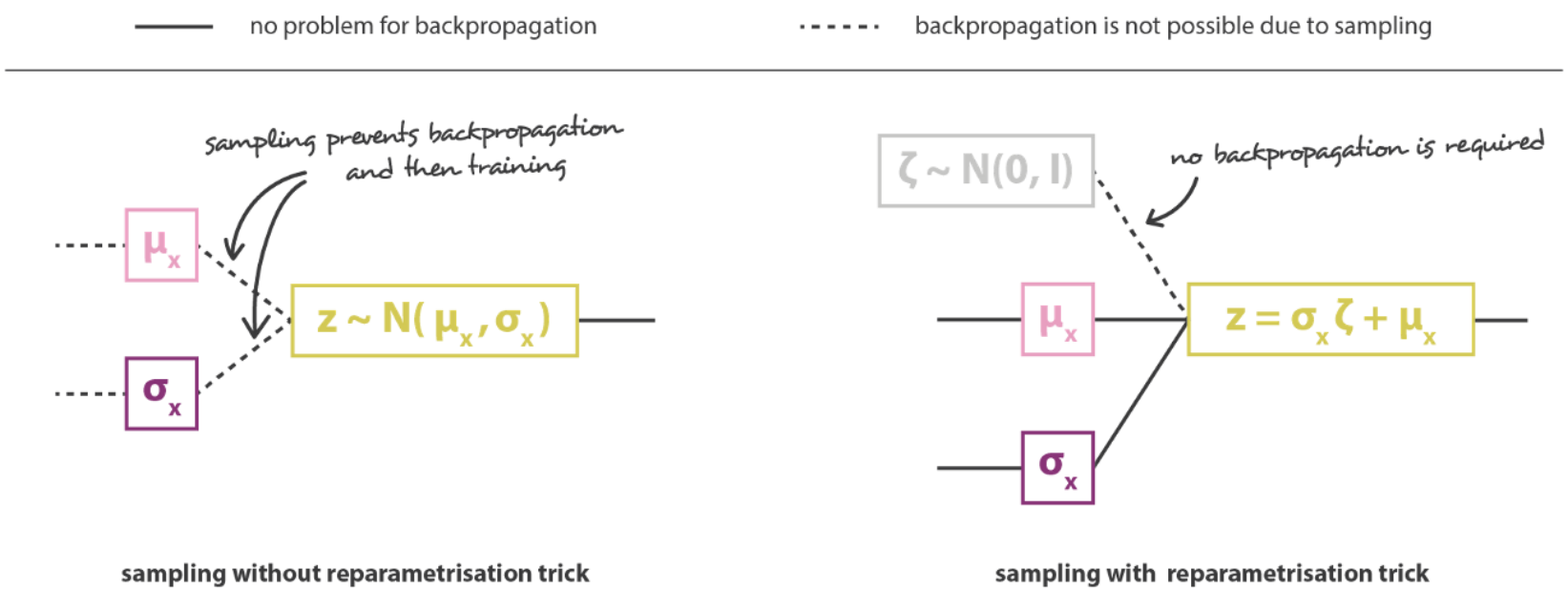


Image from [TDS – Understanding VAEs]

Solution: reparameterization trick

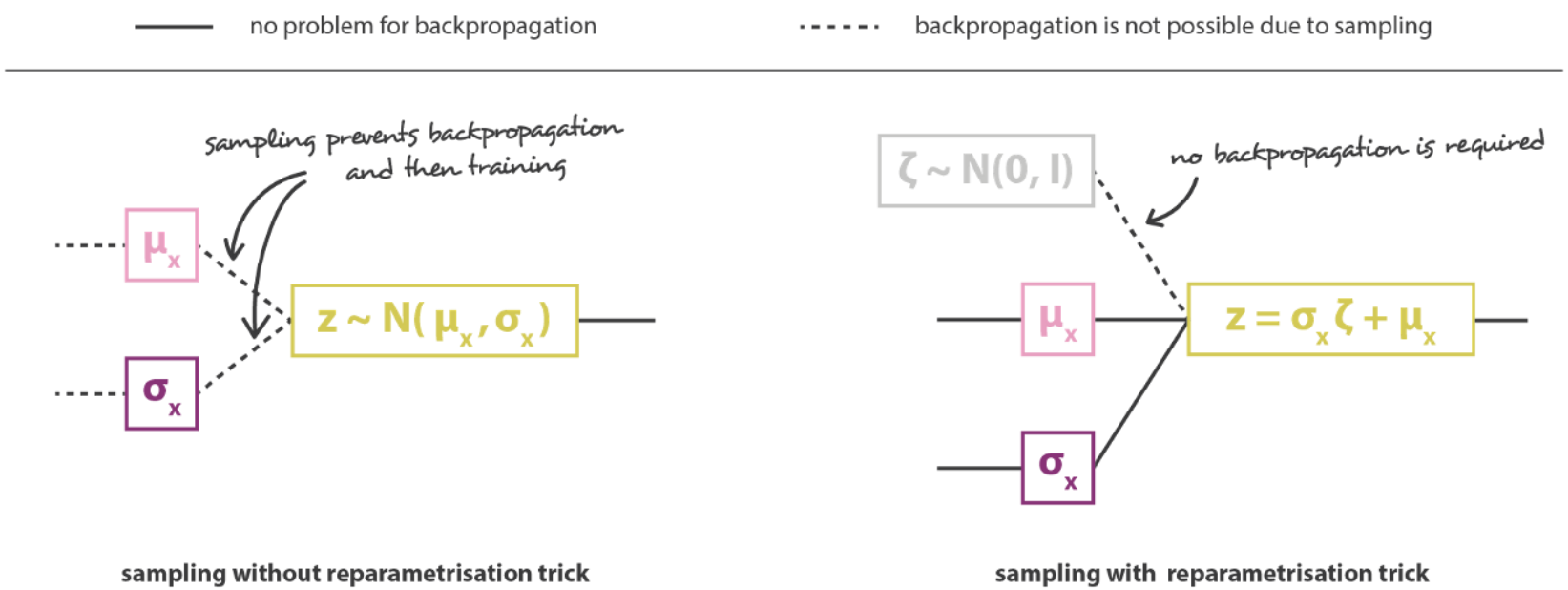
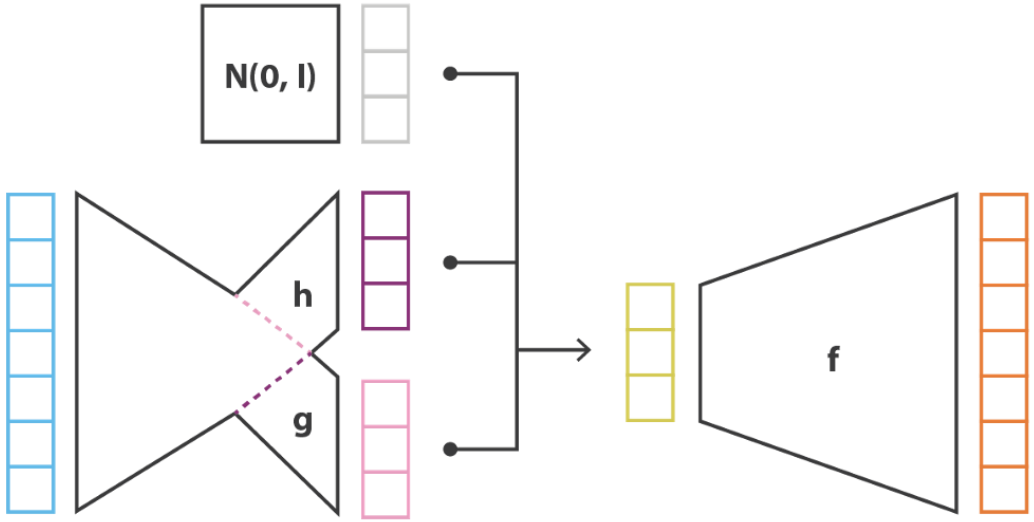


Image from [TDS – Understanding VAEs]

VAE model



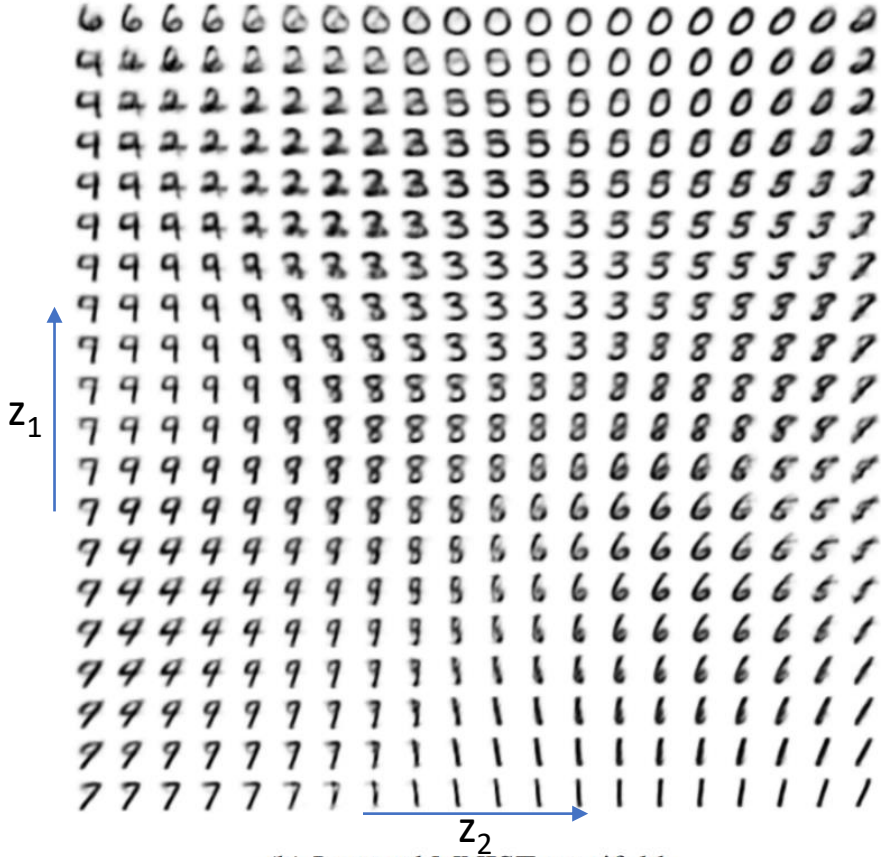
$$loss = C \| x - \hat{x} \|^2 + KL[\, N(\mu_x, \sigma_x), N(0, I) \,]$$

Image from [TDS – Understanding VAEs]

Generating data

Examples with 2-D latent space

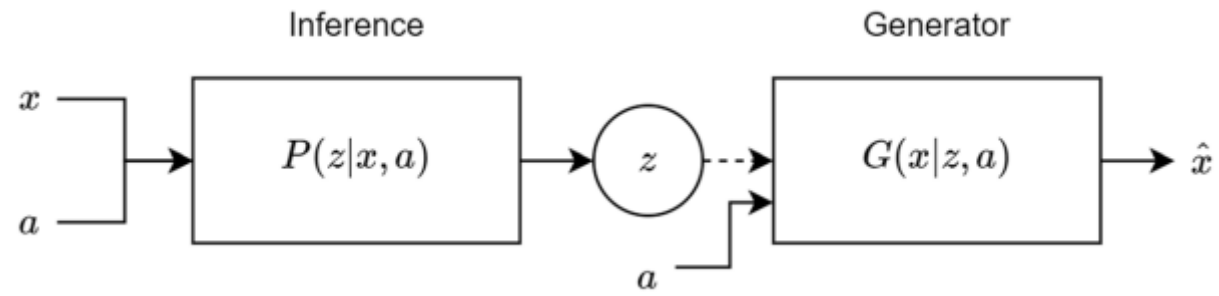
(a) Learned Frey Face manifold



(b) Learned MNIST manifold

Image from [Kingma and Welling, ICLR 2014]

Generating new proteins



AEVPSGEQLFNSNCSACHIGGNNVIISHKTLRKEALEKYAMNSLEAIRYQVVNGKNAMPAFGGRLNEEEIDAIATYVLGQAELD

ADLEAGEQIFSANCAACHGGGNNIIMPEKTLKKDALEENGMKSVEAITYQVTNGKNAMPAFGGRLSDEDIEDVANYVLSQAEKGW
ADLEHGAQIFSANCAACHAGGNNVIMPDKTLKKDALEKNGMNSIEAITYQVTNGKNAMPAFGGRLSDEDIEDVANYVLSQAEKGW
ADLENGGKVFSGACAACHIGGENIVRPEKTLKKDALEEGGMDSIEAITAQVTNGKNAAPAFGERLVDEDIEDVAEYVL
ADLAAGEQIFSANCAACHAGGNNVVMPDKTLKKDALEKYGMNSIEAITTQVTNGKNAMPAFGGRLEAEDIEDVAAYVLSQAEG
ADLEHGEQIFSANCAACHAGGNNVIMPEKTLKKDALEKYGMNSVEAITTQVTNGKNAMPAFGGRLEDEQIEDVANYVLSQSEW
ADIEHGEKIFSANCAACHAGGNNAIMRNKTLKKEALEPNGMNSIEAITYQVTNGKNAMPAFGGRLSDEDIEDVANYVLKQAEKGW
ADLAAGEQIFSANCAACHAGGNNIIMPEKTLKKEALEKYSMNSIEAITTQVTNGKNAAPAFGGRLSDEDIEDVANYVLSQAEKGW
ADIITGEQIFSANCAACHIGGNNAIRPEKTLKKPALETNGMNSVDAITTQVVNPKNAMPAFGGRLEDEDIEDVANYVLSQAEK

Image from [Greener et al, Scientific Reports 2018]

Generating new proteins

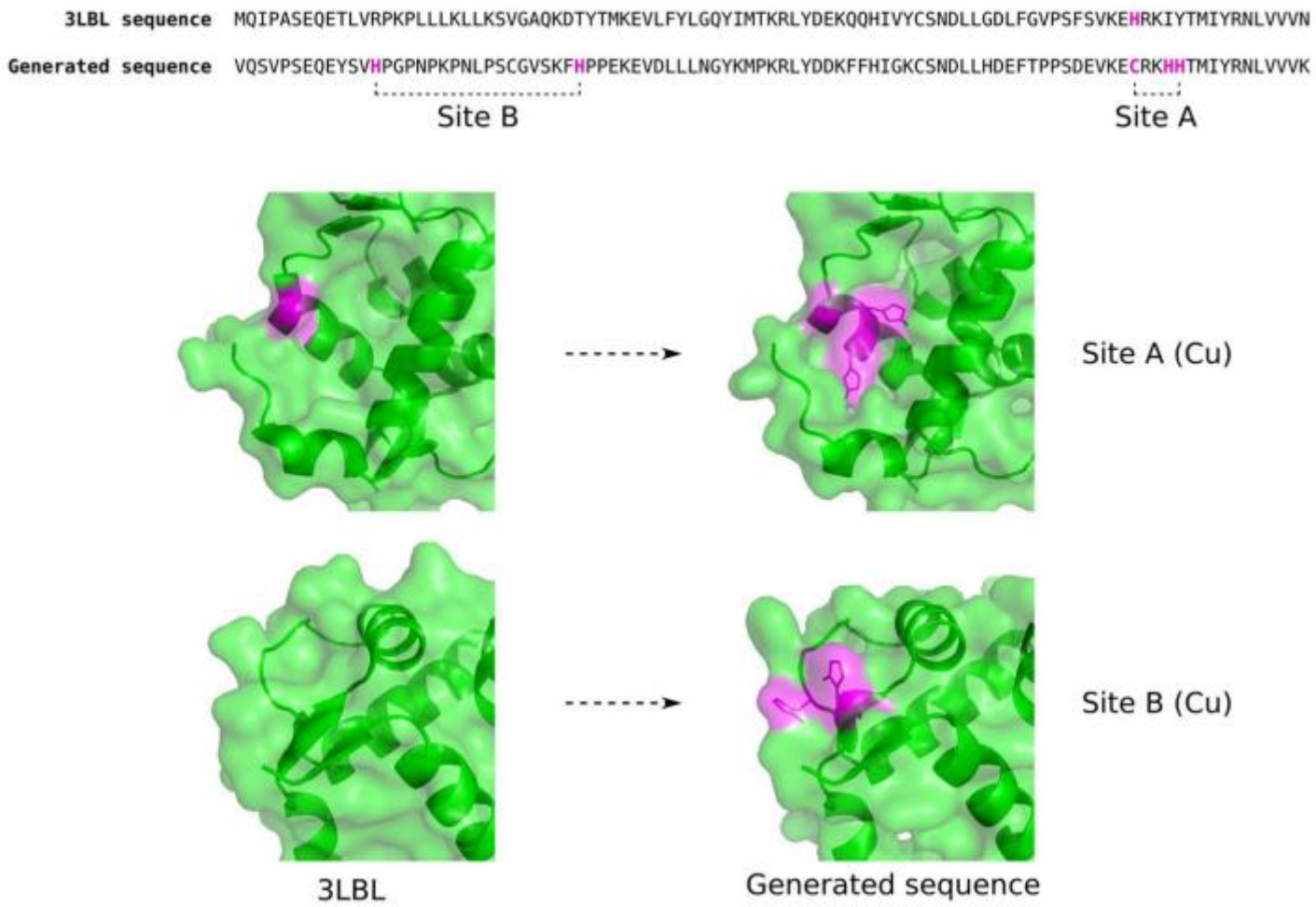


Image from [Greener et al, Scientific Reports 2018]

VAE takeaways

- Similar architecture to autoencoders
- Uses regularization (mean and variance) to ensure regular latent space
- Samples probability distribution from encoded inputs
- “Variational inference” used to derive loss function
- Used to generate new data that is similar but not equivalent to input data
- Output can be interpolations between inputs