From Autoencoders to Variational Autoencoders: The Loss Function

Valerio Velardo

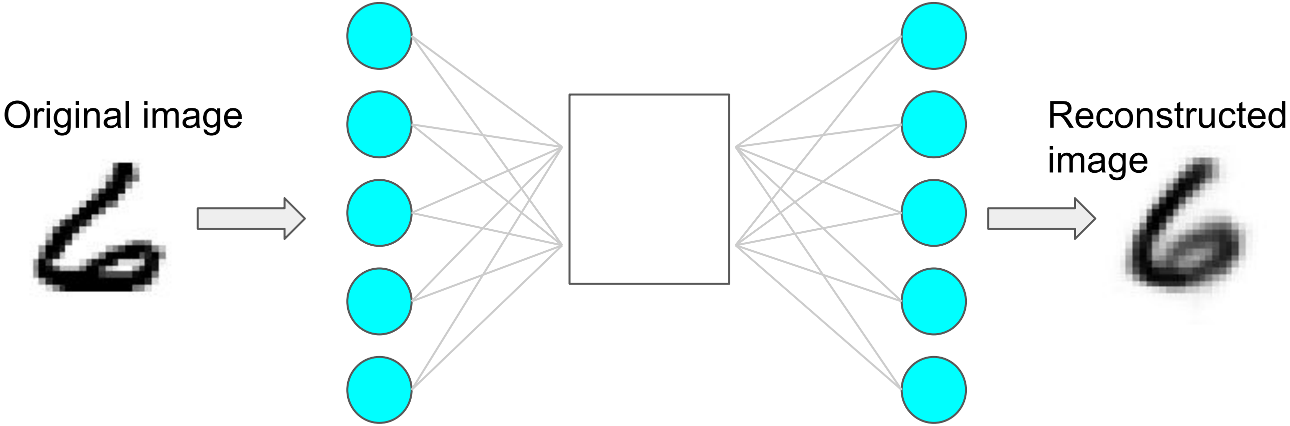
From Autoencoders to Variational Autoencoders

- Modify encoder component
- Modify loss function

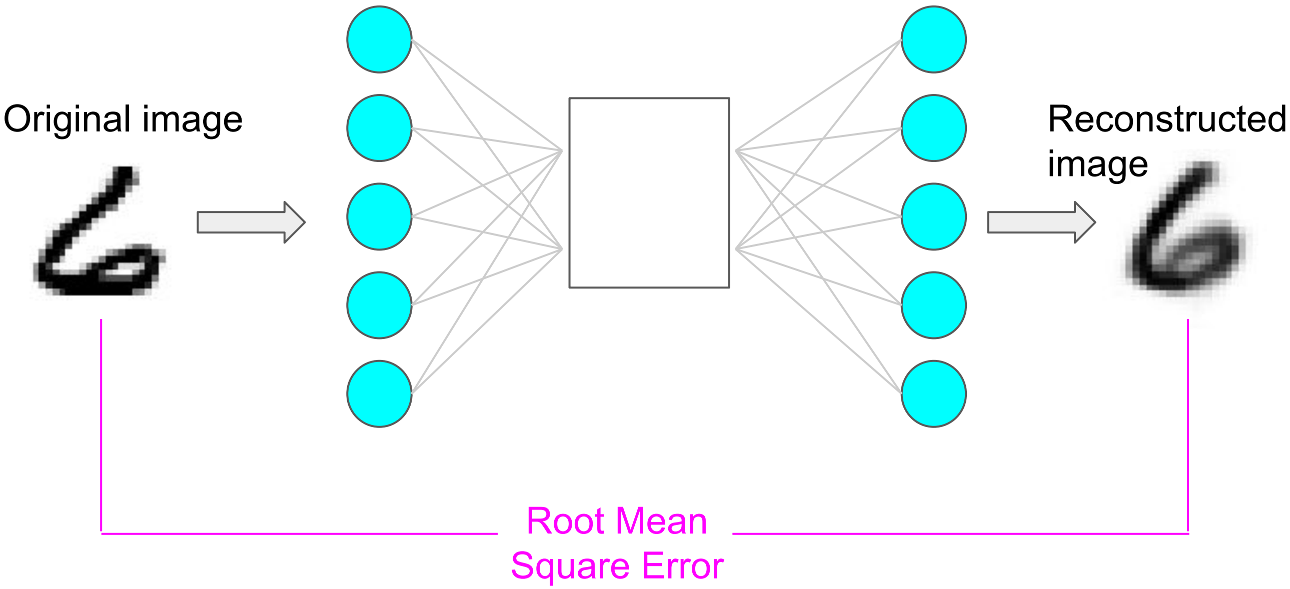
From Autoencoders to Variational Autoencoders

- Modify encoder component
- Modify loss function

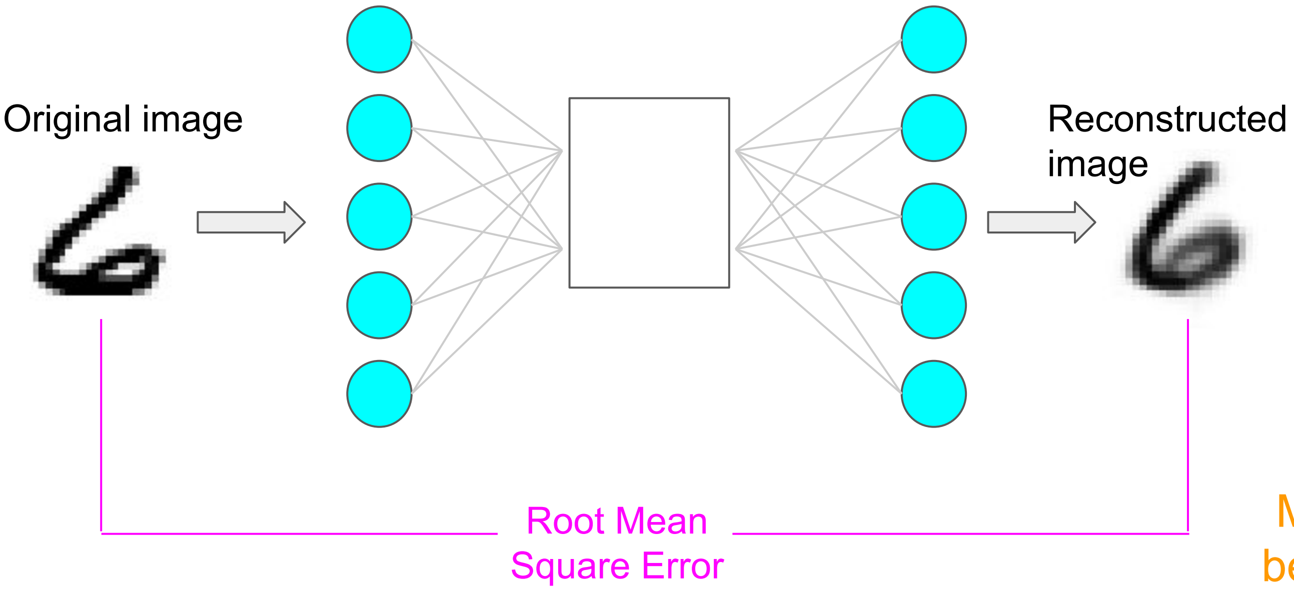
Loss function: Autoencoder



Loss function: Autoencoder



Loss function: Autoencoder



Training goal:
Minimise difference
between original and
reconstructed image

Loss function: Autoencoder

$$LOSS = RMSE$$

Loss function: Variational Autoencoder

Kullback–Leibler Divergence

$$LOSS = RMSE + \boxed{KL}$$

Loss function: Variational Autoencoder

Kullback–Leibler Divergence

$$LOSS = \boxed{RMSE} + \boxed{KL}$$

Reconstruction error

Loss function: Variational Autoencoder

Kullback–Leibler Divergence

$$LOSS = \underbrace{\boxed{RMSE}}_{\text{Reconstruction error}} + \boxed{KL}$$

Reconstruction error

Difference between normal distribution (mean vector, log variance vector) from standard normal distribution

Kullback–Leibler Divergence: The intuition

- Measures the difference between two probability distributions

Kullback–Leibler Divergence: The intuition

- Measures the difference between two probability distributions
- It's not a *distance*
  - KL divergence isn't symmetric
  - KL(A, B) ≠ KL(B, A)

Kullback–Leibler Divergence: The intuition

- Measures the difference between two probability distributions
- It's not a *distance*
  - KL divergence isn't symmetric
  - $KL(A, B) \neq KL(B, A)$
- Can be given in "closed" form with normal distributions

Kullback–Leibler Divergence (Closed form)

$$D_{KL}(N(\mu,\sigma)||N(0,1)) = \frac{1}{2}\sum(1 + log(\sigma^2) - \mu^2 - \sigma^2)$$

Kullback–Leibler Divergence (Closed form)

$$D_{KL}(\boxed{N(\mu,\sigma)}||N(0,1)) = \frac{1}{2}\sum(1 + log(\sigma^2) - \mu^2 - \sigma^2)$$

normal distro

Kullback–Leibler Divergence (Closed form)

standard normal distro

$$D_{KL}(\boxed{N(\mu,\sigma)}||\boxed{N(0,1)}) = \frac{1}{2}\sum(1 + log(\sigma^2) - \mu^2 - \sigma^2)$$

normal distro

Kullback–Leibler Divergence (Closed form)

standard normal distro

$$D_{KL}(\boxed{N(\mu, \sigma)} || \boxed{N(0, 1)}) = \frac{1}{2} \boxed{\sum} (1 + log(\sigma^2) - \mu^2 - \sigma^2)$$

normal distro

Sum across all dimensions of the latent space

Kullback–Leibler Divergence (Closed form)

standard normal distro

$$D_{KL}(\boxed{N(\mu,\sigma)} \,||\, \boxed{N(0,1)}) = \frac{1}{2}\boxed{\sum}\boxed{(1 + log(\sigma^2) - \mu^2 - \sigma^2)}$$

normal distro

Sum across all dimensions of the latent space

Loss function: Variational Autoencoder

Kullback–Leibler Divergence

$$LOSS = \boxed{RMSE} + \boxed{KL}$$

Reconstruction error

Difference between normal distribution (mean vector, log variance vector) from standard normal distribution

What does the KL loss term do?

Penalize observations where mean and log variance vectors differ significantly from the parameters of a standard normal distribution (mean vector = 0 and log variance = 0)

KL divergence loss term fixes

- Promotes symmetry around the origin
- Decreases chance of large gaps between clusters of points

Weighting the loss function

$$LOSS = \boxed{\alpha} \cdot RMSE + KL$$

reconstruction loss weight

Weighting the loss function

$$LOSS = \alpha \cdot RMSE + KL$$

- Finding the correct value for α is a delicate exercise

Weighting the loss function

$$LOSS = \alpha \cdot RMSE + KL$$

- Finding the correct value for α is a delicate exercise
- If α is too small -> poorly reconstructed images
- If α is too large -> same issues as with AE

Weighting the loss function

$$LOSS = \alpha \cdot RMSE + KL$$

- Finding the correct value for α is a delicate exercise
- If α is too small -> poorly reconstructed images
- If α is too large -> same issues as with AE
- α can be treated as a hyper-parameter to optimise

What next?

- Implement VAE