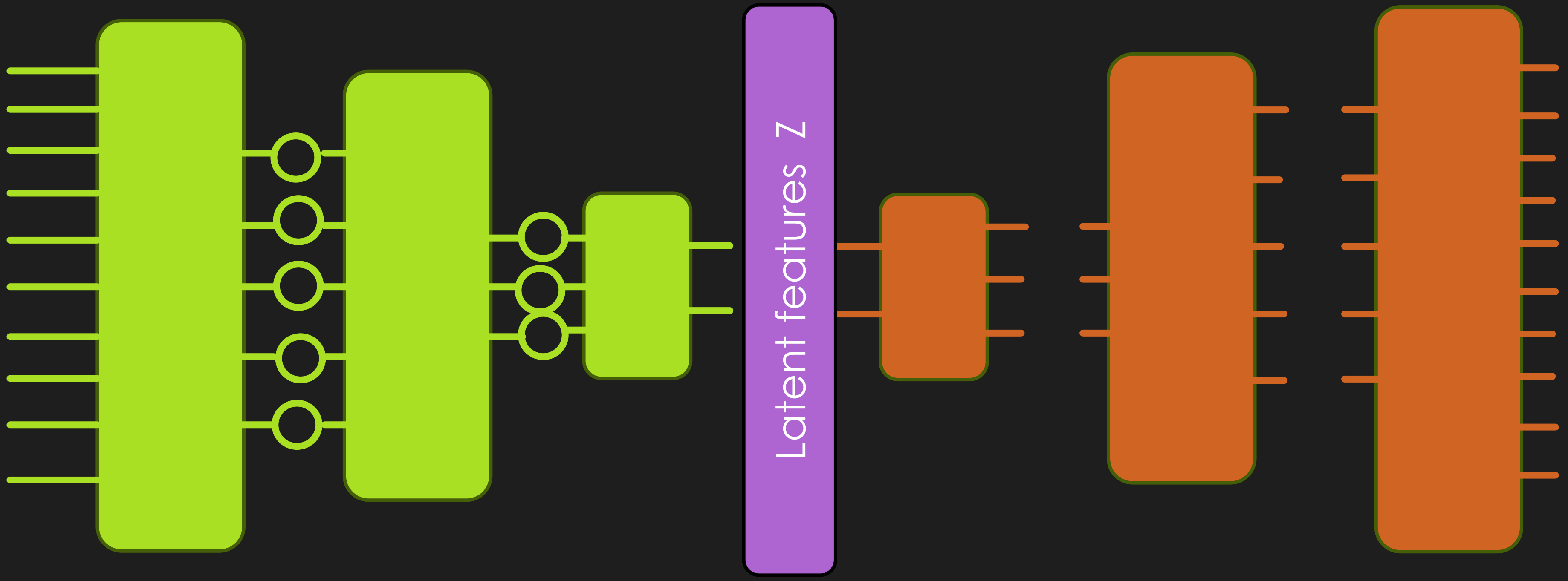
INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

Lecture 26

Hassan Ashtiani

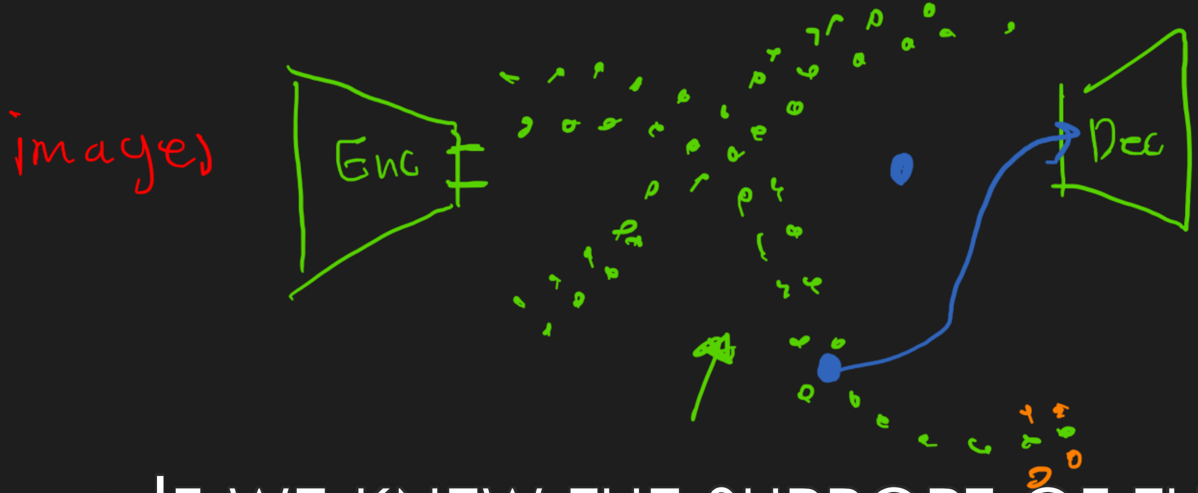
AUTOENCODER

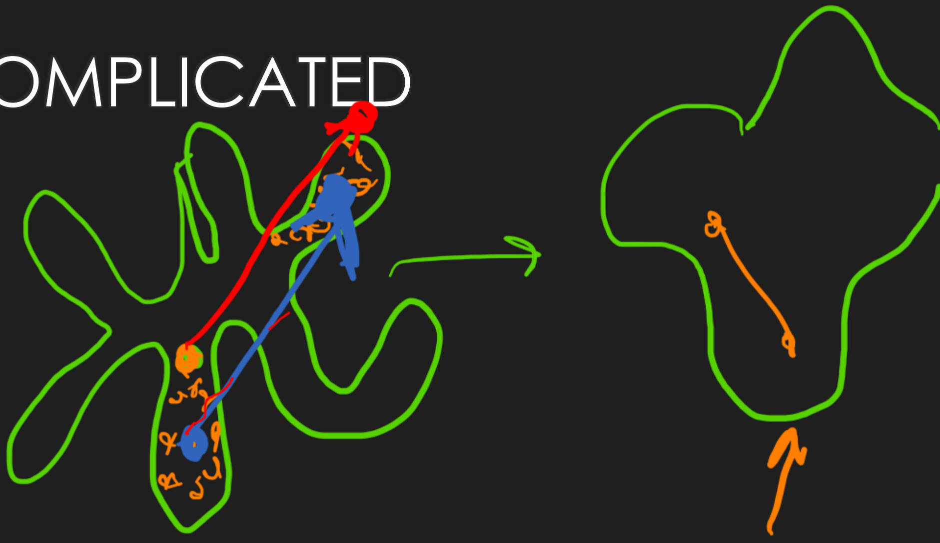


- Generate a new image that is not in the dataset?

AUTOENCODER – LATENT SPACE

- LATENT SPACE DISTRIBUTION CAN BE COMPLICATED



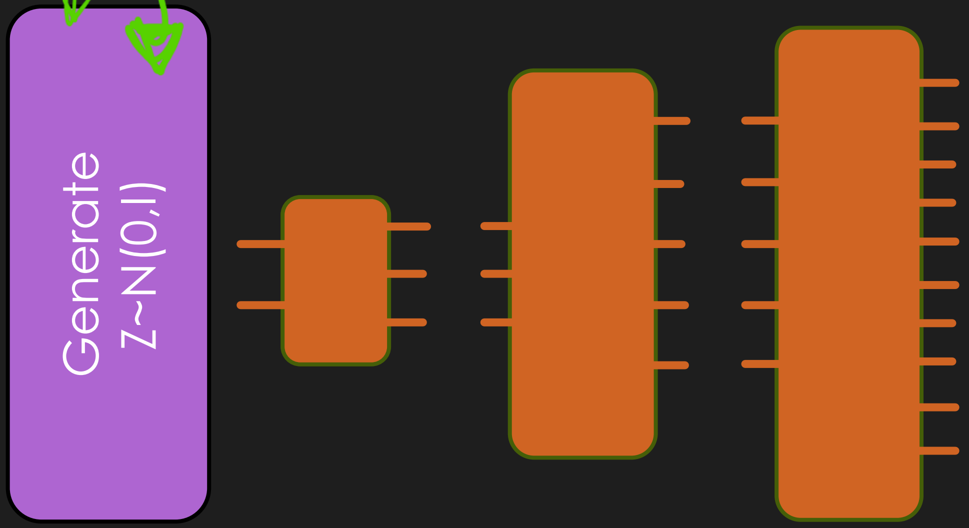


- IF WE KNEW THE SUPPORT OF THE DISTRIBUTION WE COULD HAVE GENERATED NEW SAMPLES
  - ADDING NOISE TO THE LATENT SPACE MIGHT HELP A BIT BUT..

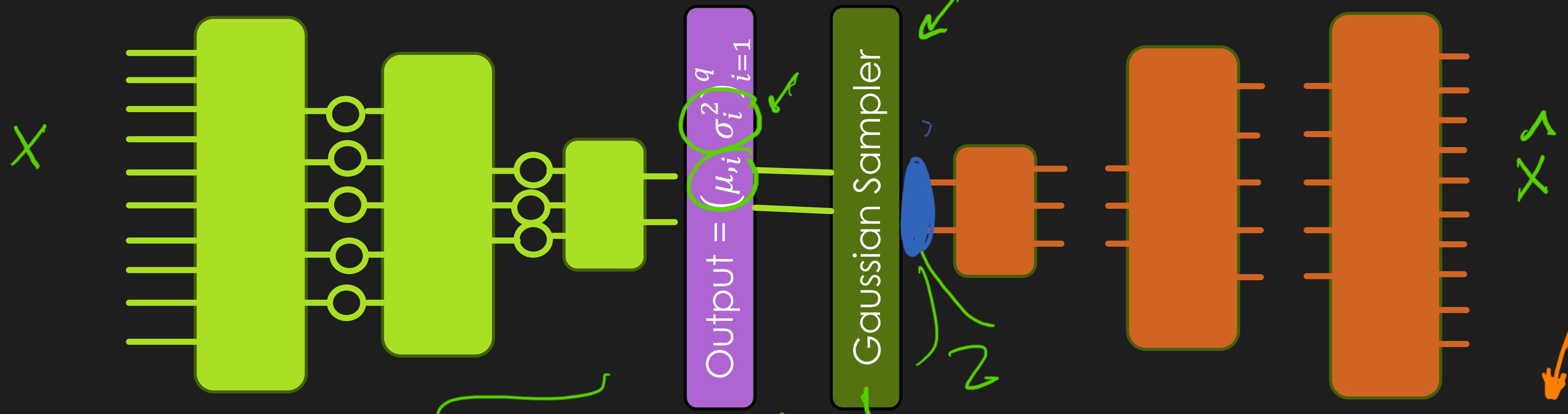
$\Sigma = I$

TAMING AUTOENCODERS

- IDEA: TRAIN AUTOENCODER IN A WAY THAT THE LATENT SPACE DISTRIBUTION LOOKS LIKE ISOTOPIC GAUSSIAN(?!)
  - DECODER LEARNS TO TURN GAUSSIAN NOISE INTO NEW IMAGES
- FOR GENERATING NEW IMAGES, SIMPLY FEED GAUSSIAN NOISE TO THE DECODER



TAMING AUTOENCODERS



- The encoder outputs μ_i, $\log \sigma_i^2$
- The sampler generate a Gaussian point $\sim N(\mu, \Sigma)$
  - $\mu = [\mu_1, \dots, \mu_q]^T$, $\Sigma = diag(\sigma_1^2, \dots \sigma_q^2)$
  - A bit similar to adding noise latent features (but noise scale is learned)
- The decoder turns this sample into an image

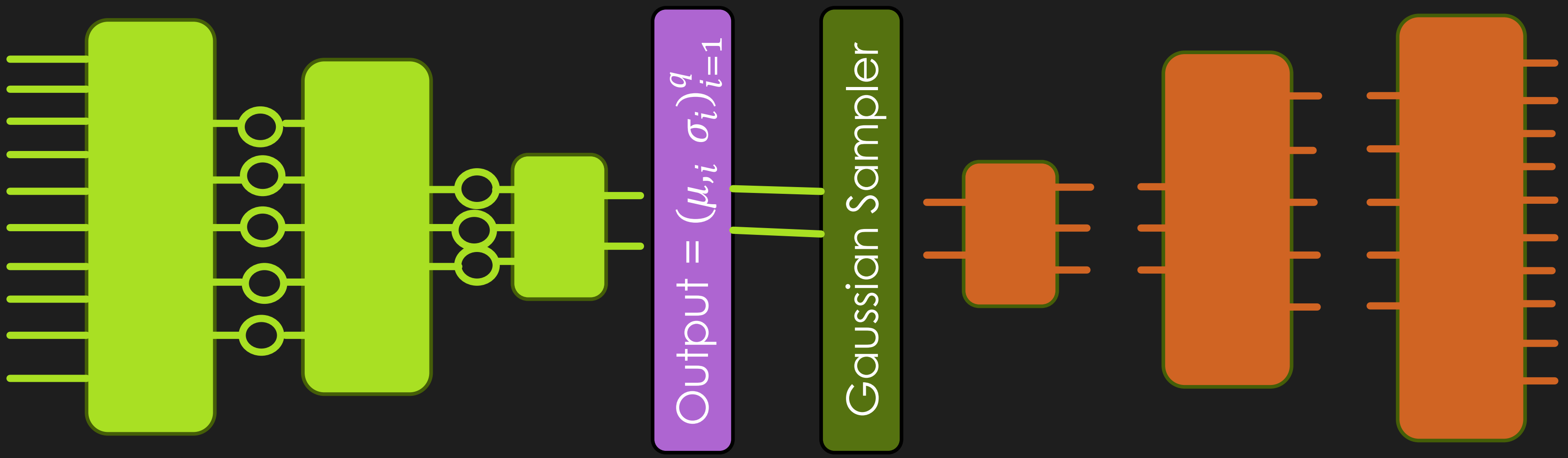
$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_q^2 \end{bmatrix}$

$z \sim N(\mu, \Sigma)$

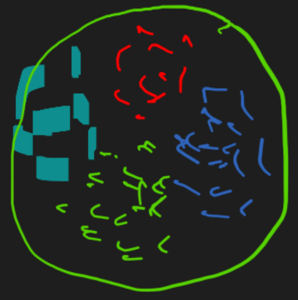
$z \sim N(\mu, I)$

$z = \mu + N(0, I)$

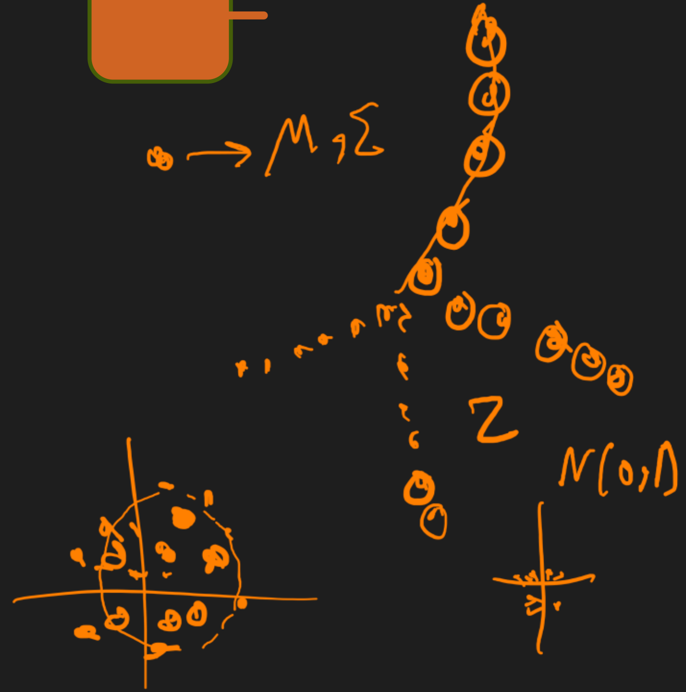
TAMING AUTOENCODERS



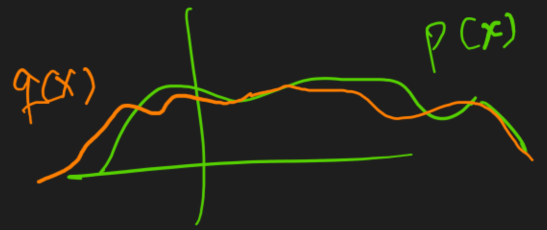
- TRAINING OBJECTIVE:
  - SMALL RECONSTRUCTION ERROR
    - E.G., $\|Dec(Sampler(Enc(x))) - x\|_2^2$ ✓
  - REGULARIZE
    - MAKE μ_i'S CLOSE TO 0 AND σ_i^2'S (CLOSE TO 1)
    - HOW?



p(z|x)
p(z)
(Z|X)
Z X
→ μ, Σ



KL DIVERGENCE



- $KL(p||q) = E_{x \sim p}\left(\mathrm{LOG}\,\dfrac{p(x)}{q(x)}\right)$

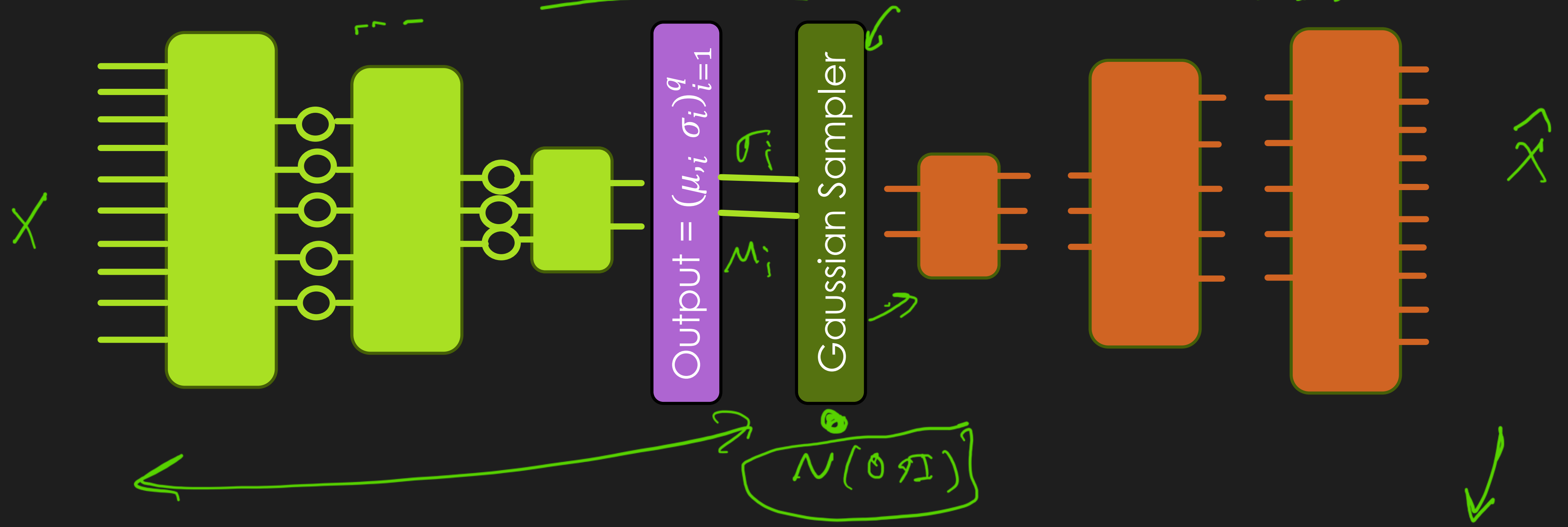
$$= \int_X p(x) \log \frac{p(x)}{q(x)}\, dx$$

- $KL(p||q) = 0 \Leftrightarrow p = q$
- $KL \geq 0$
- $KL(p||q) \neq KL(q||p)$

KL DIVERGENCE

- $KL(P||Q) = E_{x\sim p}\left(\log\frac{p(x)}{q(x)}\right)$
- Univariate Gaussians
  - $KL(N(\mu_1,\sigma_1^2)\|N(\mu_2,\sigma_2^2)) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2+(\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$
  - $KL(N(\mu_1,\sigma_1^2)\|N(0,1)) = \frac{1}{2}(-\log(\sigma_1^2) + \sigma_1^2 + \mu_1^2 - 1)$
- Uncorrelated Gaussians
  - $\mu = [\mu_1, \dots, \mu_q]^T, \Sigma = diag(\sigma_1^2, \dots \sigma_q^2)$
  - $KL(N(\mu,\Sigma)\|N(0,I)) = \sum_{i=1}^{q}\frac{1}{2}\left(-\log\left(\sigma_i^2\right) + \sigma_i^2 + \mu_i^2 - 1\right)$

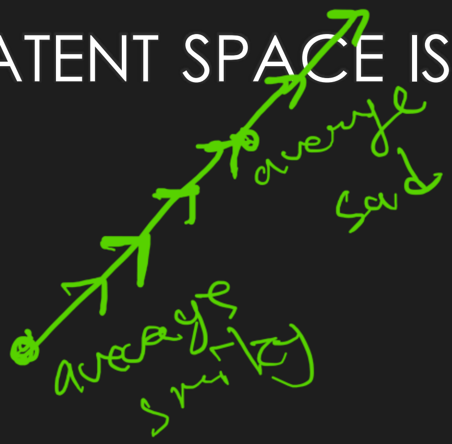
SIMPLIFIED VARIATIONAL AUTOENCODER



$$\text{MIN}\left(\underbrace{\sum_{x}\|Dec(Sampler(Enc(x))) - x\|_2^2}_{\text{Reconstruction error}} + \lambda \underbrace{\sum_{i=1}^{q}\left(-\text{LOG}\left(\sigma_i^2\right) + \sigma_i^2 + \mu_i^2\right)}_{\text{Regularizer}}\right)$$

VARIATIONAL AUTOENCODERS

- VAE's original formulation is more complicated
  - Based on evidence lower bound [ELBO], which is not covered in this course
- The latent space is sometimes interpretable



- The images are sometimes blurry
  - "GANs" can generated sharper images