

## Week 10

↳ Midterms

↳ Anonymous eval!

- generally, reflect effort in class (OH + study sessions + psets)

- goal is learning:

↳ redo one long question for up to 1/2 pts lost

↳ I'll replace 1st midterm grade w/ 2nd, if better

- let's chart in terms!

↳ Guideline of 12 hours per class per week

2.5	1.5	2	2	2	2
class	prep attempt	OH	study session	attempt	
	• reading #1 • pset			#2	

## Beyond Supervised Learning

Q : Lots of unlabelled data (images, text),  
how do we use it?

This week, we will explore pretraining

## Auto encoders

Idea: Use data as its own label

$$f_0: \mathbb{R}^d \rightarrow \mathbb{R}^k$$

"Encoder"

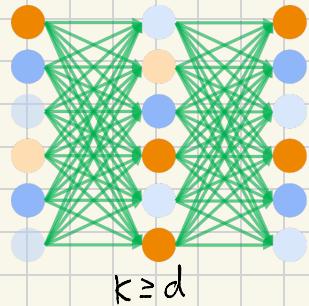
$$f_1: \mathbb{R}^k \rightarrow \mathbb{R}^d$$

"Decoder"

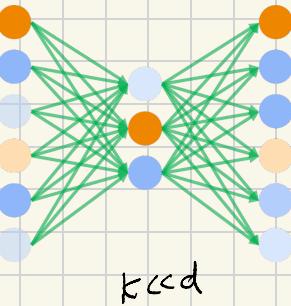
$z = f_0(x)$  is latent representation

$\tilde{x} = f_1(z)$  is reconstruction

$$\mathcal{L}_{\text{recon}} = \frac{1}{n} \sum_{i=1}^n \|x^{(i)} - \tilde{x}^{(i)}\|_2^2$$



vs.

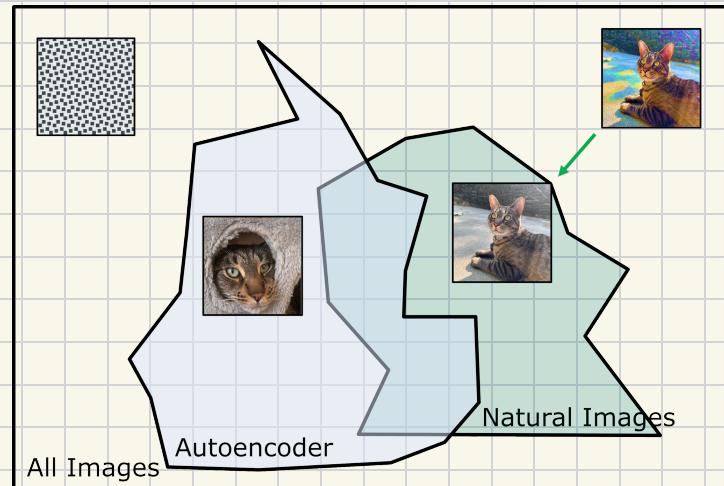


Applications include:

- data compression
- denoising
- inpainting
- representation learning

## Data Manifold

Necessarily losing information, okay because we only want to represent meaningful images



Prefer working with latent rep.

1. more efficient in lower dim
2. most meaningful features

Examples include:

- classification
- clustering
- generation

## Variational Autoencoders

Goal: make latent space "nicely behaved"

Treat latent as random sample

$$z \sim N(\mu_x, \Sigma_x)$$

↑ mean      ↗ variance  
from encoder      from decoder

"Nicely behaved"  $\Leftrightarrow N(\mu_x, \Sigma_x) \approx N(0, I)$

Distribution  $P \approx$  Distribution  $Q$

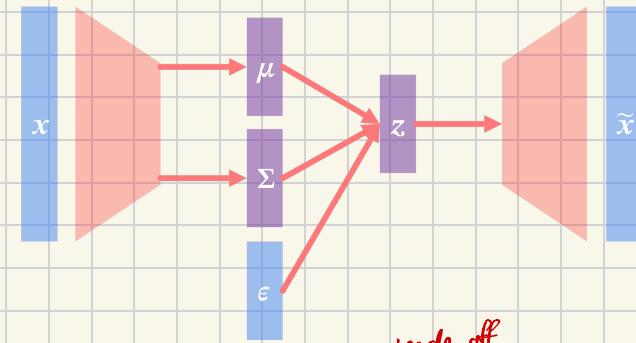
Loss: KL divergence

$$D_{KL}(P||Q) = E_{z \sim P} \left[ \log \frac{P(z)}{Q(z)} \right]$$

Example in  $\mathbb{R}$ :  $P = N(\mu_A, I)$ ,  $Q = N(\mu_B, I)$

How do backpropagate through a sample?

$$z = \mu_x + \sum_x^{1/2} \epsilon \in N(0, I) \text{ draw}$$



$$\mathcal{L}_{VAE} = \mathcal{L}_{\text{recon}} + \lambda D_{KL}(P||Q)$$

$\uparrow$   
 $z$  encodes  
meaning of  $x$

$\uparrow$   
 $z$  distributed  
similarly

trade off

Anonymous eval!