Tutorial on Variational Autoencoders

Variational Autoencoders (VAEs)

↳ circa 2016 emerged as on the most popular approaches to unsupervised

built ontop of standard function approxim's (neural networks)

type of "generative" modelling

goal of generative models is to learn some distribution $P(x)$ that is as close as possible to $P_{gt}(x)$ (ground truth)

VAE

↳ Assumptions are weak (this is good)

↳ Backprop train is "fast"

1.1 Preliminaries: Latent variable models

w/ gen models, the more complicated the dependancies between dimensions = harder to train

A gen model performs well when it decides which character [0-9] to draw before determining any pixels

↳ this is latent variables

2 Variational Autoencoders

Mathematically, VAEs have little to do w/ classical autoencoders

trad autoencoder has encoder - decoder

In training VAEs resemble this structure

to solve train equation VAEs have 2 problems to solve

① how to define latent variables, z (decide what info they hold)

② how to deal w/ integral over z

eq = $P(x) = \int P(X \mid z; \theta)\, P(z)\, dz$

① latent vars

for digits need to not just encode the digit in its latent dimensions but its abstract stylistic properties z

we want to avoid deciding by hand the info each dimension of z encodes & the dimensions dependancies

VAEs assume the is no simple interp of the dimensions of z

instead they assert that ~~these~~ samples of z can be drawn from a simple dist
→ $N(0, I)$
I = identity matrix

how?

A distribution in d dimensions can be generated by taking a set of d variable w/ a norm distributed & mapping them through a sufficiently complicated function

example:
→ norm dist points of 2D graph
→ eg: $g(z) = z/10 + z/\|z\|$
→ ring shape of same points on 2D graph

provided powerful function approx (NN) we can simply learn a function which maps our indep, norm dist z values to whatever latent variables are needed for the model

↳ then map latent vars to X

$N(X \mid f(z, \theta))$
→ f can be an NN

within the NN layers = ① learn latent
② map to pixels to render

dont need to hand craft the latent structure - or make sure it exists

if there exists a latent structure that maximizes the likelihood (ability to reprod) then the NN layers will uncover it