- WGAN - MNIST example
- ATLAS Fast calogan example
  - Michele's slides

---

## Next generative model framework: Variational autoencoders

- example of a latent variable model

data $x \longrightarrow$ latent variable $z$

encodes "meaning" of $x$ (aka embedding)
ideally in a much reduced, simpler space

To sample:

$z \sim p(z)$ ← "prior" some fixed, simple dist'n

$x \sim p_\theta(x|z)$ This gives $(x,z) \sim p(x,z)$

Throw away $z \to x \sim p(x)$.

Objective: max likelihood

WANT: $p(x) = p_{data}(x)$.

$$\max_\theta \sum_{x \sim p_{data}} \log p_\theta(x) = \max_\theta \sum_{x \sim data} \log \int p_\theta(x|z) p(z) dz$$

↓ direct eval via MC integ?

$$\max_\theta \sum_{x \sim data} \log \frac{1}{N} \sum_{z \sim p(z)} p_\theta(x|z)$$

- Computationally expensive
- Noisy gradients
- Curse of dimensionality (MC needs large N in higher dim)

→ direct MC generally fails!

Instead: variational approach + tractable lower bound on likelihood

introduce $q_\phi(z|x)$ ← proposal dist'n for more efficient sampling than prior

$$\log p(x) = \log \int dz \; p_\theta(x|z)\,p(z)\,\frac{q_\phi(z|x)}{q_\phi(z|x)}$$

$$\approx \log \frac{1}{N} \sum_{z \sim q_\phi(z|x)} \frac{p_\theta(x|z)\,p(z)}{q_\phi(z|x)}$$

→ sampling would be most efficient if $q_\phi(z|x) \sim p_\theta(x|z)\,p(z)$

i.e. integrand $O(1)$ always

So $q_\phi$ wants to be posterior!

more to come...

"Jensen ineq"

$\log \frac{1}{N}\sum f \geq \frac{1}{N}\sum \log f$

(AKA AM-GM ineq)

$$\geq \frac{1}{N} \sum_{z \sim q_\phi(z|x)} \log \frac{p_\theta(x|z)\,p(z)}{q_\phi(z|x)}$$

"evidence lower bound"

ELBO

thing inside log wants to be $p(x)$

Bayesian evidence

The ELBO is tractable!

$$\max_{\theta,\phi} \sum_{x \sim p_{data}(x)} \text{ELBO} = \max_{\theta,\phi} \sum_{\substack{x \sim p_{data}(x) \\ z \sim q_\phi(z|x)}} \log \frac{p_\theta(x|z)\,p(z)}{q_\phi(z|x)}$$

Just train w/ batches of $x$ and $z$ together

Another look at the ELBO:

$$KL(P\|Q) = \int dx\, P(x) \log \frac{P(x)}{Q(x)}$$

"Kullback Leibler divergence" → aka "relative entropy"

not a metric distance, asymmetric $P \leftrightarrow Q$

non negative $KL \geq 0$ but not bounded

$0$ iff $P=Q$

- $JSD(P,Q) = \frac{KL(P\|M) + KL(Q\|M)}{2}$ $M = \frac{P+Q}{2}$.

Introduce $p(z|x) = \frac{p_\theta(x|z)\, p(z)}{p(x)}$ true posterior

$$ELBO = \frac{1}{N} \sum_{z \sim q_\phi(z|x)} \log \frac{p(z|x)\, p(x)}{q_\phi(z|x)}$$

$$= \boxed{\log p(x) - KL(q_\phi(z|x) \| p(z|x))} \leq \log p(x)$$

So ELBO is saturated when $q_\phi$ is the true posterior ✓.

confirming our intuition → "variational Bayesian inference"

can use to infer posteriors

↓

So what is our generative model?

need $p_\theta(x|z)$ and $q_\phi(z|x)$ (w/ $z \sim p(z)$ fixed prior)

"probabilistic decoder" "probabilistic encoder"

$x \to z \to x$ "variational autoencoder"

Can choose many things for $p_\theta$ & $q_\phi$ (NNs)

e.g. $\begin{cases} q_\phi(z|x) = \mathcal{N}(\mu_\phi(x), \sigma_\phi(x)) \\ p_\theta(x|z) = \mathcal{N}(\mu_\theta(z), \beta_\theta(z)) \end{cases}$ normal dist'n w/ mean/var DNNs ①

Very simple ansatz...

Another common choice: continuous Bernoulli dist'n

$p(x|z) = N(z) z^x (1-z)^{1-x}$

$x \in [0,1]$

$z \in [0,1]$

Cont. gen. of discrete Bern.

$x = 0$ or $1$

$p(x|z) = z \quad x=1$
$\qquad\quad\; 1-z \quad x=0$

or BCE for cont. Bern

$x \log z + (1-x)\log(1-z)$

often $N(z)$ ignored but not strictly correct...

---

Another look at ELBO loss

$$-\text{ELBO} = \underbrace{-\mathbb{E}_{z \sim q_\phi(z|x)} \log p_\theta(x|z)}_{} + \underbrace{KL(q_\phi(z|x) \| p(z))}_{\text{regularizer}}$$

max → min

$$\mathbb{E}_{z \sim q_\phi(z|x)} \frac{\|x - \mu_\theta(z)\|^2}{2\beta_\theta(z)^2}$$

"reconstruction error"

want $x \to z \to x$ to recover $x$

standard Autoencoder loss

"regularization"

promotes $x \to z \sim p(z)$

regular latent space

---

"Reparam trick": if $q_\phi$ & $p(z)$ are Gaussian, can show

$$KL(q_\phi | p(z)) = \frac{1}{2}\left(-1 + \mu_\phi(x)^2 + \sigma_\phi(x)^2 - \log \sigma_\phi(x)^2\right)$$

So Gaussian VAE objective becomes finally:

$$\min_{\theta,\phi} \sum_{\substack{x \sim \text{data} \\ z \sim q_\phi(z|x)}} \left[ \frac{\|x - \mu_\theta(z)\|^2}{2\beta_\theta(z)^2} + \frac{1}{2}\left(-1 + \mu_\phi(x)^2 + \sigma_\phi(x)^2 - \log \sigma_\phi(x)^2\right) \right]$$

In practice, allowing $\beta_\theta(z) \to$ unstable traj.

$\to \beta_\theta = \beta$ const.

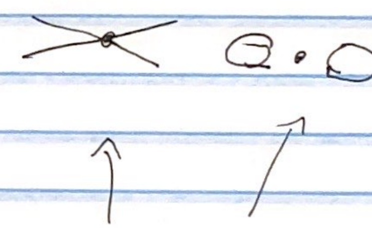
"$\beta$-VAE"

Like vanilla AE + regularizer term

& probabilistic/generative interpretation

---

- MNIST example
- radio astronomy example



2102.01007

Bastien, Scaife et al

Trained VAE to generate FRI and FRII radio galaxy images

↳ ~ only ~600 images (!!)

+ augmentation (eg random rotations)

$\to$ 200k images

Also CVAE

conditioned on label

- SDSS example 2002.10464

Galaxy spectra