WGAN loss

\[
L(x) = \max_G \min_{\|\nabla \phi\|_\infty} \mathbb{E}_{x \sim p_{\text{data}}} \phi(x) - \mathbb{E}_{x \sim p_{\text{gen}}} \phi(x)
\]

Enforce Lipschitz: many approaches

- original paper: "weight clipping"

\[ w \rightarrow \text{clip}(w-c) \quad \text{if } w > c \]
\[ w \rightarrow w \quad \text{if } w \leq c \]

This enforces \( \mathbb{K} \)-Lipschitz crudely.

\[ f(x) = g_2(g_1(\ldots g_1(x) \ldots)) \]

So \[ \nabla f(x) \sim [\nabla g_1(\ldots g_1(x) \ldots)] \]

So if weights bounded, \( \nabla f \) bounded.

Better way: gradient penalty \( \mathbb{E}_{y \sim p_{\text{data}}} \| \nabla_2 f(x; y) \|^2 \]

add regularizer to loss \( \mathcal{L} = \mathcal{L} + \lambda \mathbb{E}_{y \sim p_{\text{data}}} \| \nabla_2 f(x; y) \|^2 \)

P: dist of pts \( x = \beta x + (1-\beta) y \)

Some theory behind this...

\[ y \rightarrow \text{dist gen and } z \]

not clear why seems to work \( \mathbb{E} y = 1 \)

instead \( \mathbb{E} y = 1 \) ... but empirically works
- WGAN - MNIST example
- ATLAS Fast cegan example
- Michele’s notes

Next generative model framework: Variational autoencoders

- example of a latent variable model
  \[ \text{data } x \rightarrow \text{ latent variable } z \]
  encodes “meaning of x” (aka embedding) ideally in a much reduced, simpler space

To sample:
  \[ z \sim p(z) \]
  “prior”
  \[ x \sim \phi(x|z) \]
  This gives \( (x,z) \sim p(x) \)
  Thus \( x \rightarrow x \sim p(x) \).

Objective: max likelihood

\[ \text{max } \sum_{x \sim \text{data}} \log p(x) = \max_{\theta} \sum_{x \sim \text{data}} \log \int p(x) p(z|x) \, dz \]

\[ \text{want: } p(x) = p(x|z) \]

\[ \text{direct eval na MC?} \]

\[ \text{may } \sum_{x \sim \text{data}} \log \frac{1}{N} \sum_{z \sim p(z)} p(x|z) \]
- Computationally expensive
- Many gradient samples
- Case of dimensionality (MC needs large N in hydridic)
- Direct MC generally fails!

Instead: variational approach + tractable lower bound on likelihood

\[
\log p(x) = \log \int d\theta \ p(x|\theta)p(\theta) \ \frac{q_\phi(\theta|x)}{p(x|\theta)p(\theta)}
\]

\[
\log \frac{q_\phi(\theta|x)}{p(x|\theta)p(\theta)} \rightarrow \text{simple and most efficient if}
\]

\[
\text{"Jensen"} \quad \frac{1}{N} \sum_{i=1}^{N} \log p(x_i|\theta)p(\theta) \geq \log \frac{1}{N} \sum_{i=1}^{N} \log p(x_i|\theta)p(\theta)
\]

\[
(\text{AKA AMPC-EM}) \quad \text{"Evidence Lowerbound"}
\]

\[
\text{ELBO} \quad \text{they inside log tractable ph}
\]

\[
\text{Bayesian evidence}
\]

The ELBO is tractable!

\[
\max_{\theta, \phi} \quad \text{ELBO} = \max_{\phi} \ \log p(x|\theta)p(\theta)
\]

\[
\text{Just train w/ batches of } x \text{ and } z \text{ together}
\]