Introduction to Score-matching

Justin T Chiu

July 3, 2023

Goals

1. What is an energy-based model and why are they hard to train?

2. What is score-matching, and how can it be used to train an EBM?

3. How does score-matching relate to diffusion models?

Energy-Based Models (EBM)

Problem setup: Density estimation

- Observations x

- Goal: Learn a model $p(x)$
  - Capture uncertainty / variability over x

- Participation: Give examples of an x we model, and how $p(x)$ is parameterized
  - Ex: Language modeling uses Transformers for $p(x) = \prod_t p(x_t | x_{<t})$

Running example: Image generation

- "Solved": Finite-class density estimation
    - Softmax assigns a score to each $E(x)$ then normalizes

$$softmax(x) = \frac{\exp(E(x))}{\sum_x \exp(E(x))}$$

- Image generation
    - Can consider every small change in a single pixel as a new clas
    - Size: 1024 x 1024, each pixel has 256 * 3 values

- More efficient paradigms

Image generation models

- Autoregressive: Break down generation from left-to-right

$$p(x) = \prod_t p(x_t | x_{<t})$$

- Latent variable model: Break down generation more flexibly

$$p(x) = \sum_z p(x|z) p(z)$$

- Energy-based model: Don't enforce breakdown of decision process

What is an EBM?

- Globally normalized over x, eg sentences

$$p(x) = \frac{\exp(E(x))}{Z}$$
$$Z = \int_x \exp(E(x))$$

- Computation of the partition function Z is hard