InfoGAN: Information Maximizing Generative Adversarial Networks

Yuancheng Yu

October 14, 2021
Motivation

- Unsupervised learning: tasks unknown at training
- Disentangled representation: explicit salient features
  - facial expression, eye color, hairstyle, eyeglasses, identity
  - MNIST: digit, angle, thickness of stroke
- hossRBM (higher-order spike-and-slab restricted Boltzmann Machines)
  - Discrete features
  - Exponential cost in number of features
- InfoGAN
  - Discrete and continuous
  - Efficient
Method

- Recall GAN:

\[ \min_G \max_D V(D, G) = \mathbb{E}_{X \sim P_{\text{data}}} [\log D(X)] + \mathbb{E}_{Z \sim P_Z} [\log(1 - D(G(Z)))] \]

- Issue: \( Z \) may be entangled
- Idea: \( G(Z, C) \)
  - "latent code" \( C = (C_1, \ldots, C_L) \), where \( P(C) = \prod_i P(C_i) \)
- Mutual information \( I(X; Y) = H(X) - H(X | Y) \)
  - amount of info learned from \( Y \) about \( X \)
- Intuition: maximize \( I(C; G(Z, C)) \)
- InfoGAN:

\[ \min_G \max_D V_I(D, G) = V(D, G) - \lambda I(C; G(Z, C)) \]
Method

- **Issue:** $I(C; G(Z, C))$ hard to maximize without $P(C | X)$
- **Idea:** Variational Information Maximization
  - Lower bound by approximating $P(C | X)$ with $Q(C | X)$

$$I(C; G(Z, C)) = H(C) - H(C | G(Z, C))$$

$$= \mathbb{E}_{X \sim G}[\mathbb{E}_{C' \sim P}[\log P(C' | X)]] + H(C)$$

$$= \mathbb{E}_{X \sim G}[D_{KL}(P \parallel Q) + \mathbb{E}_{C' \sim P}[\log Q(C' | X)]] + H(C)$$

$$\geq \mathbb{E}_{X \sim G}[\mathbb{E}_{C' \sim P}[\log Q(C' | X)]] + H(C)$$