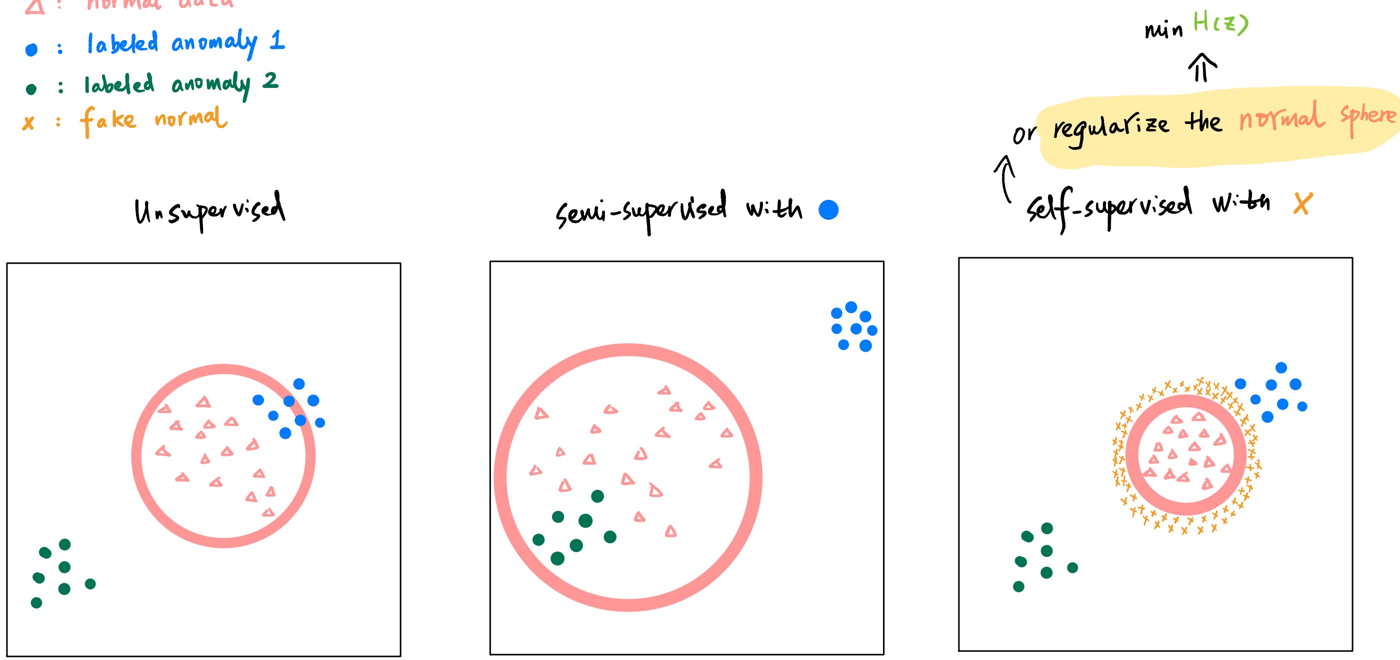
Intuition for Bias from Latent Space

- △ : normal data
- ● : labeled anomaly 1
- ● : labeled anomaly 2
- x : fake normal



A Unified Objective for Anomaly Detection [1]

Principle $\max \; KL(p_n(X,Z) \,\|\, p_a(X,Z))$

- $$= \max \; \underbrace{I_n(X,Z)}_{\text{mutual info}} - \underbrace{H_n(Z)}_{\text{entropy}} + \underbrace{\mathbb{E}_{p_n(X)}[H(p_n(Z|X), p_a(Z|X))]}_{\text{Cross-entropy}}$$

[Can we gain an "unbiased" estimate for that w/ limited skewed data?]

Lower Bound for Unsupervision

$$\geq \max \; I_n(X,Z) - H_n(Z)$$

1: Zhang et al, 2020

A Unified Objective for Anomaly Detection

Principle $\max \ KL\left(p_n(x,z) \,\|\, p_a(x,z)\right)$

- $$= \max \ \underbrace{I_n(x,z)}_{\text{mutual info}} - \underbrace{H_n(z)}_{\text{entropy}} + \underbrace{\mathbb{E}_{p_n(x)}\left[H\left(p_n(z|x), p_a(z|x)\right)\right]}_{\text{Cross-entropy}}$$

Connection to AE

$$\max \ I_n(x,z)$$

$$= \max \ H_n(x) - H_n(x|z)$$

$$= \max \ \mathbb{E}_{x \sim p_n(x)} \mathbb{E}_{z \sim p_n(z|x)} \left[\log p_n(x|z)\right]$$

reconstruction likelihood $\Rightarrow$ goal of AE

A Unified Objective for Anomaly Detection

Principle $\max \ KL(p_n(x,z) \,\|\, p_a(x,z))$

- $= \max \ \underbrace{I_n(x,z)}_{\text{mutual info}} - \underbrace{H_n(z)}_{\text{entropy}} + \underbrace{\mathbb{E}_{p_n(x)}[H(p_n(z|x), p_a(z|x))]}_{\text{Cross-entropy}}$

Connection to SVDD $\max I_n(x,z)$; then $\max -H_n(z)$

- Step 1 for $I_n(x,z)$: pretrain using autoencoder
- Step 2 for $-H_n(z)$: minimizing empirical variances.

  $H_n(z) \propto \log \sigma^2$ if z is isotropic Gaussian

  Thus, $\min H_n(z) \Rightarrow \min_{x \in \text{normal}} \|\phi(x;\theta) - c\|^2$
- Heuristic for SAD : $\max H_a(z) \Rightarrow \max_{x \in \text{anomalies}} \|\phi(x;\theta) - c\|^2$