(Deep Infomax): Contrastive Setup:

[MAP → feature vector
MXM]

input x

output y

family of encoder: $\mathcal{E}_\psi = \{E_{\psi,\varphi}\}_{\varphi \in \Psi}$

$U_{\psi,\rho}$ → distribution over encoding for input x.

Target: $y \sim E_\psi(x)$ for some input

① maximize: $I(X, E_\psi(x))$ (MI)

② constraint: $U_{\psi,\rho}$ match V → another prior distribution

infomax

[Figure 3] mother of all:

(global vs local)

|| All

take one line

So, positive & Negative Samples:

wis 1 (11)

MI estimation & maximization

$$I_w(x;y) : E_J[T_w(x,y)] - \log E_M[e^{T_w(x,y)}]$$ // DV

joint — marginal

$T_w \rightarrow$ Discriminator.

so maximization

$$(\hat{w}, \hat{\psi})_G : \arg\max_{w,\psi} \hat{I}_w(x; E_\psi(x))$$

$\mapsto$ low label features.

$$E_\psi : f_\psi \circ C_\psi$$

$$T_{w,\psi} = D_w \circ g \circ (C_\psi, E_\psi)$$

Instead using JSD,

InfoNCE

$$\hat{I}_{w,\psi}(x, E_\psi(x)) := E_P\left[T_{\psi,w}(x, E_\psi(x)) - E_{\tilde{P}}\left[\log \sum_{x'} e^{T_{\psi,w}(x', E(x))}\right]\right]$$

This paper

needs to be high — should be low

JSD

$$\hat{I}_{w,\psi}(x, E_\psi(x)) = E_P[-sp(-T_{\psi,w}(x, E_\psi(x)))] - E_{P\times\tilde{P}}[sp(T_{\psi,w}(x', E(x)))]$$

only maximize not really the precise value

$$sp(z) = \log(1 + e^z)$$

(~~requires smaller negative sample~~)

better than NCE

Local MI maximization

feature map

$$C_\psi(x) : \{c_\psi^{(i)}\}_{i=1}^{H \times W}$$

$$E_\psi(x) = f_\psi \circ C_\psi(x)$$

$$(\hat{w}, \hat{\psi})_L = \underset{\psi, w}{\arg\max} \frac{1}{M^2} \sum_{i=1}^{M} \hat{I}_{w,\psi}\left(C_\psi^{(i)}(x); E_\psi(x)\right)$$

matching representation with prior

$$(\hat{w}, \hat{\psi})_P = \underset{\psi}{\arg\min} \underset{\phi}{\arg\max} \hat{D}_\phi\left(V \| U_{\psi,P}\right)$$

$$= E_V\left[\log D_\phi(y)\right] + E_P\left[\log\left(1 - D_\phi\left(E_\psi(x)\right)\right)\right]$$

$\Downarrow$

Adversarial Auto encoder.

final loss = weighted sum of the previous three