contrastive predictive Coding:-

$$I(x;c) = \sum_{x,c} p(x,c) \log \frac{p(x|c)}{p(x)}$$

data

context

$$z_t = g_{enc}(x_t)$$

(non linear, Shared encoder)

g_{AR} summarizes the autoregressive model (means context)

$$c = g_{AR}(z_{\le t})$$

insted of prediction, predict density ratio.

$$f_k(x_{t+k}, c_t) \propto \frac{p(x_{t+k}|c_t)}{p(x_{t+k})} \quad \text{unnormalized.}$$

→ observation from (t+k) time.

$$f_k(x_{t+k}, c_t) = \exp(z_{t+k}^T W_k c_t) \quad // \text{ single value}$$

Depends on step.

info NCE loss:- $$\mathcal{L}_N = -\mathbb{E}_X \left[\log \frac{\overbrace{f_k(x_{t+k}, c_t)}^{max}}{\sum_{x_j \in X} \underbrace{f_k(x_j, c_t)}_{minimize}}\right]$$

$$P(d=i|X, c_t) = \frac{p(x_i|c_t)\prod_{l\ne i} p(x_l)}{\sum_{j=1}^{N} p(x_j|c_t)\prod_{l\ne j} p(x_l)}$$

$$= \frac{p(x_i|c_t)/p(x_i)}{\sum_{j=1}^{N} \frac{p(x_j|c_t)}{p(x_j)}} \propto f(x_{t+k}|c_t)$$

optimal Solution bayesian (if well trained!)

$$\boxed{I(x_{t+k}, c_t) \ge \log N - \mathcal{L}_N}$$