$$\mathcal{L}_{NCE} = -\mathbb{E} \log \frac{e^{f(x_1;y)}}{\frac{1}{N}\sum_{n=1}^{N} e^{f(x_n;y)}} \qquad f: \mathbb{X} \times \mathbb{X} \to \mathbb{R}$$

$$I(x_1;y) \geqslant -\mathcal{L}_{NCE}$$

1)

$$I(x_1;y) = KL(p(x,y) \| p(x)p(y)) =$$

$$= \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \, dx\, dy =$$

$$= \int p(x,y) \log \frac{p(x|y)}{p(x)} \, dx\, dy =$$

$$= \{ q(x|y) - \text{arbitrary} \} = \int p(x,y) \log \frac{p(x|y)\, q(x|y)}{p(x)\, q(x|y)} \, dx\, dy$$

$$= \underbrace{\int p(x,y) \log \frac{p(x|y)}{q(x|y)} \, dx\, dy}_{} + \underbrace{\int p(x,y) \log q(x|y)\, dx\, dy}_{} -$$

$$\underbrace{- \int \log p(x) \underbrace{\left[\int p(x,y)\, dy \right]}_{p(x)} dx}_{} =$$

$$= \int p(x|y)\, p(y) \log \frac{p(x|y)}{q(x|y)} \, dx\, dy + \mathbb{E}_{p(x,y)} \log q(x|y)$$

$$+ H(x) =$$

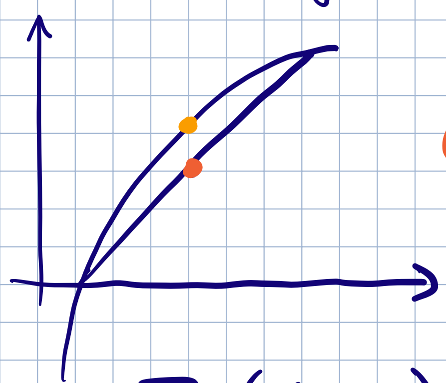
$$= \mathbb{E}_{p(y)} KL\left(p(x|y) \,\|\, q(x|y)\right) + \mathbb{E}_{p(x,y)} \log q(x|y) + H(x)$$

$$\geq \mathbb{E}_{p(x,y)}\left[\log q(x|y)\right] + H(x) =: I_{LB}$$

2) $$q(x|y) = \frac{1}{Z(y)} p(x) e^{g(x,y)}; \quad Z(y) = \underbrace{\int p(x) e^{g(x,y)} dx}$$

$$I_{LB} = \mathbb{E}_{p(x,y)}\left[\cancel{\log p(x)} + g(x,y) - \log Z(y)\right] + \cancel{H(x)}$$

$$= \mathbb{E}_{p(x;y)}\left[g(x,y)\right] - \mathbb{E}_{p(y)}\left[\log \underbrace{Z(y)}\right]$$



$$\mathbb{E} \log \xi \leq \log \mathbb{E} \xi$$

$$I(X_1; Y) \geq \mathbb{E}_{p(X_1, y)}\, g(X_1; Y) - \mathbb{E}_{p(y)} \log Z(y)$$

$$\|$$

$$\mathbb{E}_{p(X_{2:N})} I(X_1; Y) \geq \mathbb{E}_{p(X_1,y)p(X_{2:N})}\, g(X_1; X_{2:N}; Y) - \mathbb{E}_{p(y)p(X_{2:N})} \log Z(y; X_{2:N})$$

{Jensen's inequality}

$$\geq \underbrace{\mathbb{E}_{p(X_1,y)p(X_{2:N})}\, g(X_1; X_{2:N}; Y)} - \log \mathbb{E}_{p(y)p(X_{2:N})} Z(y; X_{2:N})$$

$$X_1; Y \perp X_2, \ldots, X_N$$

$$g(x_{1:N}; y) = \log \frac{e^{f(x_1; y)}}{\sum_{n=1}^{N} e^{f(x_n; y)}}$$

$$I(x_1; y) \geq -\mathcal{L}_{NCE} \overbrace{- \log \underbrace{\mathbb{E}_{p(y)p(x_{2:N})} z(y; x_{2:N})}}^{\log N}$$

$$\mathbb{E}_{p(y)p(x_{2:N})} z(y; x_{2:N}) = \mathbb{E}_{p(y)\underbrace{p(x_{2:N})p(x_1)}} \frac{e^{f(x_1; y)}}{\sum_{n=1}^{N} e^{f(x_n; y)}} =$$

$$= \mathbb{E}_{p(y)p(x_{1:N})} \frac{e^{f(x_1; y)}}{\sum_{n=1}^{N} e^{f(x_n; y)}} = \sum_{i=1}^{N} \frac{1}{N} \mathbb{E}_{p(y)p(x_{1:N})} \frac{e^{f(x_i; y)}}{\sum_{n=1}^{N} e^{f(x_n; y)}}$$

$$= \frac{1}{N} \mathbb{E}_{\underbrace{p(y)p(x_{1:N})}} \frac{\cancel{\sum_{i=1}^{N} e^{f(x_i; y)}}}{\cancel{\sum_{n=1}^{N} e^{f(x_n; y)}}} = \frac{1}{N}$$

$$q(x|y) \propto p(x) e^{g(x,y)}$$

$$\parallel$$

$$p(x|y)$$

$$e^{g(x,y)} \propto \frac{p(x|y)}{p(x)}$$

$$\underbrace{\frac{e^{\underline{f(x_1; y)}}}{\sum_{n=1}^{N} e^{f(x_n; y)}}} \propto \frac{p(x|y)}{p(x)}$$