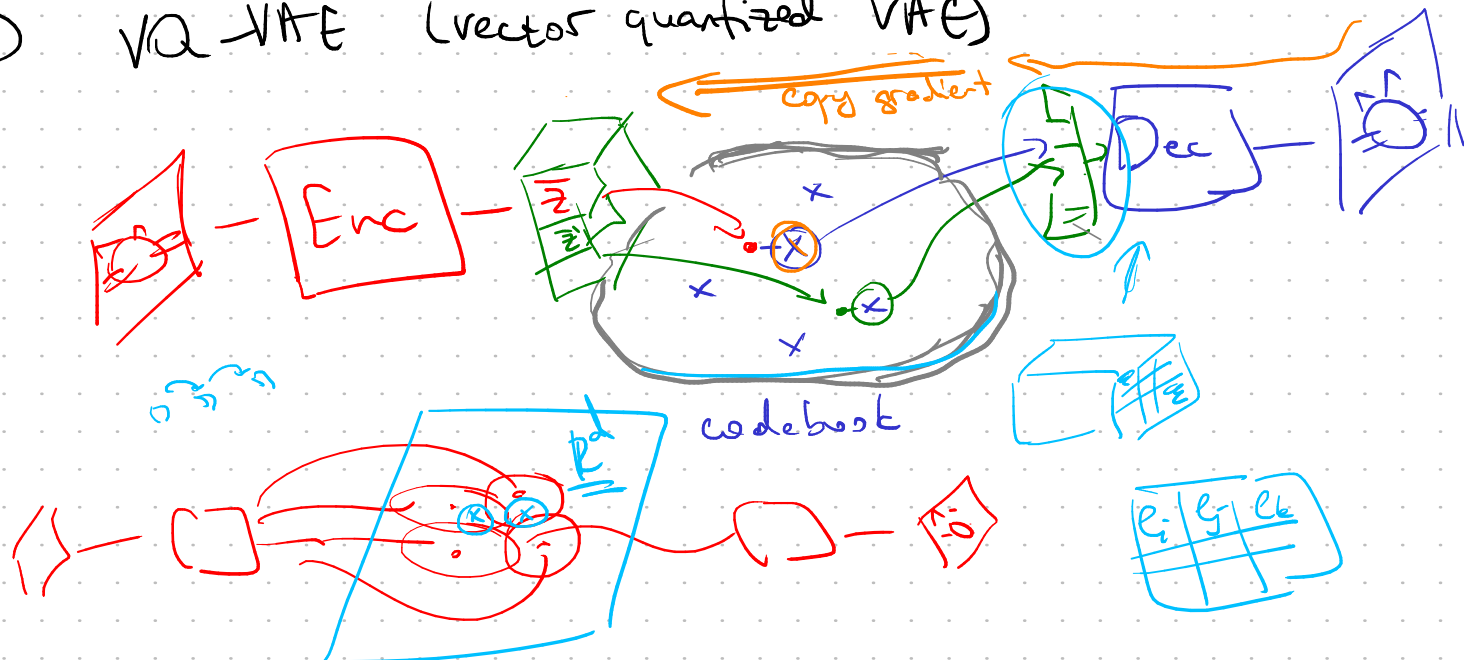
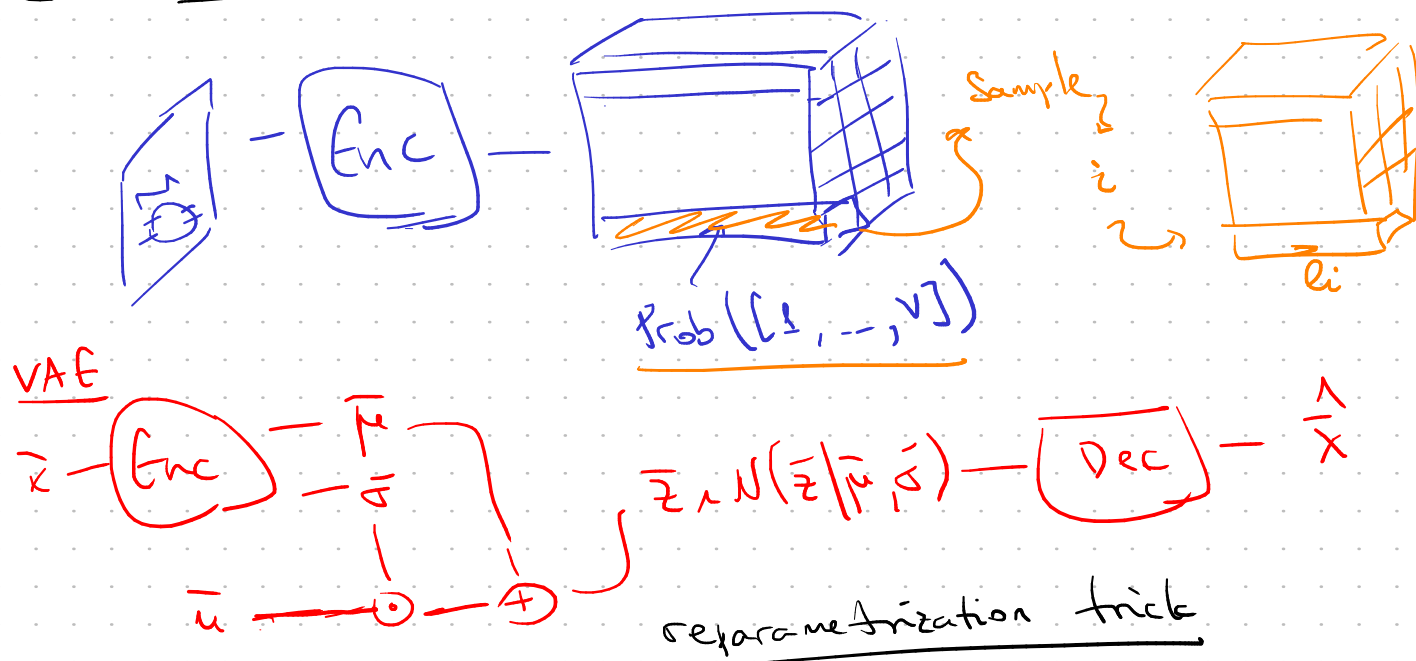


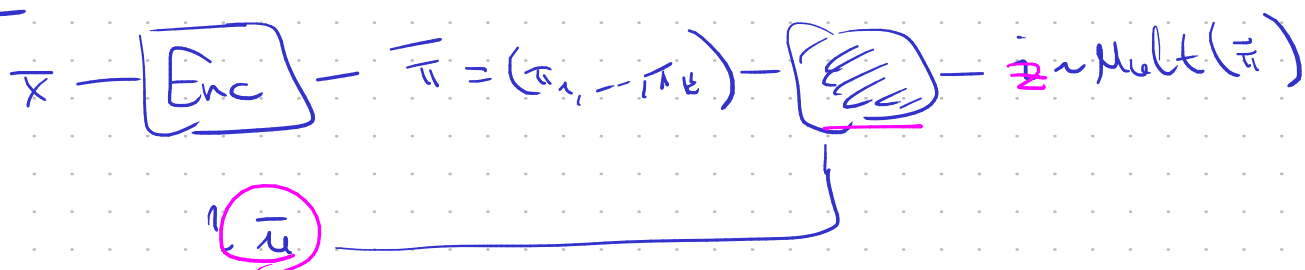
① VQ-VAE (vector quantized VAE)



② dVAE - distributional VAE



dVAE



Gumbel - Max trick

$$z \sim \text{Mult}(\bar{\pi}) \Leftrightarrow \underline{z} = \underset{i}{\text{argmax}} (g_i + \log \pi_i), \text{ use } \bar{\pi} = (\pi_1, \dots, \pi_K)$$

$$g_i \sim \text{Gumbel} \quad p(g_i) = e^{-(g_i + e^{-g_i})}, \quad F(g_i) = e^{-e^{-g_i}}$$

Proof.
$$p(z=k) = p(\forall j, g_k + \log \tau_k \geq \underline{g_j} + \log \tau_j) =$$
$$= \int_{-\infty}^{\infty} p(\forall j \neq k, \underline{g_j} + \log \tau_j \leq g_k + \log \tau_k) p(g_k) dg_k =$$

$$= \int_{-\infty}^{\infty} \prod_{j \neq k} p(g_j \leq g_k + \log T_k - \log T_j) \underbrace{p(g_k)}_{= F(g_k + \log T_k - \log T_j)} dg_k =$$

$$= \int_{-\infty}^{\infty} \prod_{j \neq k} e^{-e^{-g_k - \log \pi_k + \log \pi_j}} \cdot e^{-(g_k + e^{-g_k})} dg_k =$$

$$= \int_{-\infty}^{\infty} \prod_{j \neq k} e^{-\pi_j} \cdot \overbrace{e^{-(g_k + \log \tau_k)}}^{1 - \pi_k} \cdot e^{(g_k + e^{-g_k})} dg_k =$$

$$= \int_{-\infty}^{\infty} e^{-e^{-(g_k + \log \pi_k)}} \cdot \left(\sum_{j \neq k} \pi_j \right) \cdot e^{-g_k - e^{-g_k}} dg_k =$$

$$= \int_{-\infty}^{\infty} e^{-e^{-(g_k + \log \tau_k)} + \pi_k \cdot e^{-g_k - \log \tau_k} - g_k - e^{-g_k}} dg_k =$$

$$-e^{-(g_k + \log \tau_k)} \cdot (g_k + \log \tau_k) + g_k + \log \tau_k + \tau_k \cdot e^{-g_k - \log \tau_k}$$

$$= e^{\log \bar{y}_k} \int_{-\infty}^{\infty} e^{-(g_k + \log \bar{y}_k)} dg_k = \tau_k$$

$$\bar{x} \rightarrow [\text{Enc}] \rightarrow \bar{\pi} \rightarrow [\log]$$

$$\bar{e}_2 \rightarrow [\text{Dec}] \rightarrow \hat{\bar{x}}$$

$$\bar{g} \sim \text{Gumbel} \rightarrow (+) \rightarrow \text{argmax} \rightarrow \bar{z} \sim \text{Mult}(\bar{\pi})$$

Gumbel - Softmax trick

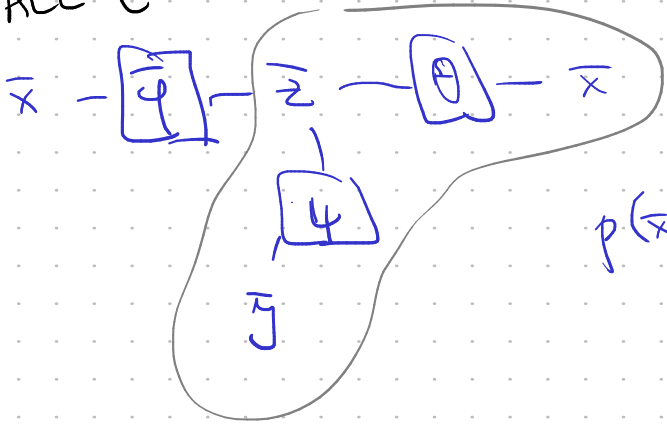
$$\bar{x} \rightarrow [\text{Enc}] \rightarrow \bar{\pi} \rightarrow [\log]$$

$$\bar{e} = \sum \alpha_i \bar{e}_i \rightarrow [\text{Dec}]$$

$$\bar{g} \rightarrow (+) \rightarrow \text{softmax} \rightarrow \bar{\alpha}$$

$$\alpha_i = \text{softmax} \left(\frac{1}{\tau} (\bar{g} + \log \bar{\pi}) \right)_i = \frac{e^{\frac{1}{\tau} (\bar{g}_i + \log \bar{\pi}_i)}}{\sum_j e^{\frac{1}{\tau} (\bar{g}_j + \log \bar{\pi}_j)}}$$

③ DALL-E

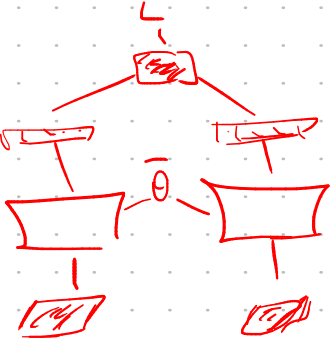


$$p(\bar{x}, \bar{y}, \bar{z}) = p_{\theta}(\bar{x} | \bar{z}) \cdot p_{\phi}(\bar{z} | \bar{y})$$

④ Neural information retrieval

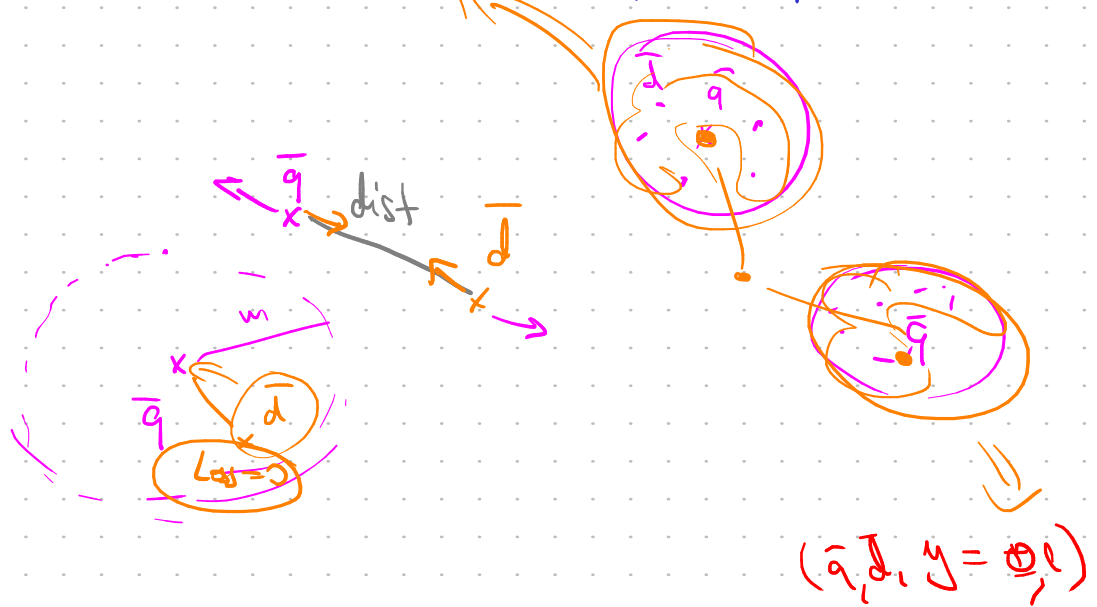
Learning to rank

Siamese networks



RankNet $\rightarrow$ LambdaRank

MART / LambdaMART



$$\mathcal{D} = \{(\bar{q}, d^+, d)\}$$

