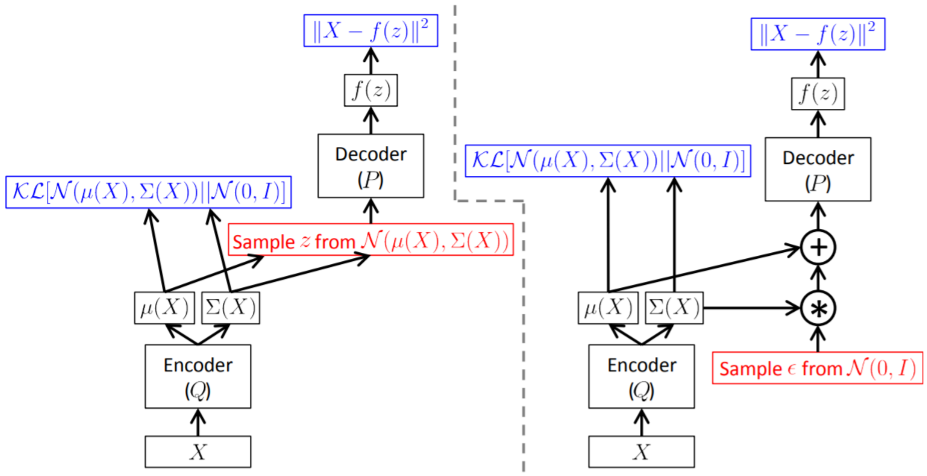
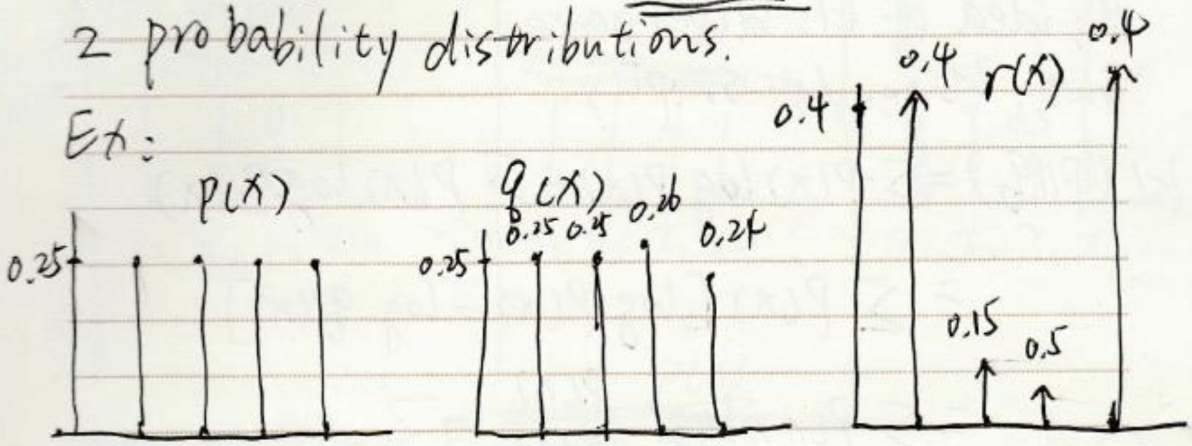


$\|X - f(z)\|^2$
$f(z)$
Decoder (P)
$\mathcal{KL}[\mathcal{N}(\mu(X), \Sigma(X))||\mathcal{N}(0, I)]$
Sample z from $\mathcal{N}(\mu(X), \Sigma(X))$
$\mu(X)$
$\Sigma(X)$
Encoder (Q)
X
$\|X - f(z)\|^2$
$f(z)$
$\mathcal{KL}[\mathcal{N}(\mu(X), \Sigma(X))||\mathcal{N}(0, I)]$
Decoder (P)
+
$\mu(X)$
$\Sigma(X)$
*
Encoder (Q)
Sample ϵ from $\mathcal{N}(0, I)$
X

KL Divergence - Relative $\xrightarrow{\text{using}}$ Entropy

1. What? Measures the distance between 2 probability distributions.

Ex:



KL(p||q) is small
KL(p||r) is large
With respect to p 意思是各用P表示

HOW: find each entropy for $P(x), q(x), r(x)$

$$\Rightarrow H_P(x) = -\sum P(x)\log P(x)$$

$$H_q(x) = -\sum q(x)\log q(x)$$

$$\Rightarrow KL(P||q) = \left[-\sum q(x)\log q(x)\right] - \left[-\sum P(x)\log P(x)\right]$$

$$= H_q - H_p$$

relative to p

$\because$ with respect to p, $\therefore KL(P||q) =$

$$\left[-\sum P(x)\log q(x)\right] - \left[-\sum P(x)\log P(x)\right]$$

$$KL(q||P) = \left[-\sum q(x)\log P(x)\right] - \left[-\sum q(x)\log q(x)\right]$$

relative to q

Although this formulation helps to explain the idea of KL divergence,
But, it needs simpliy.

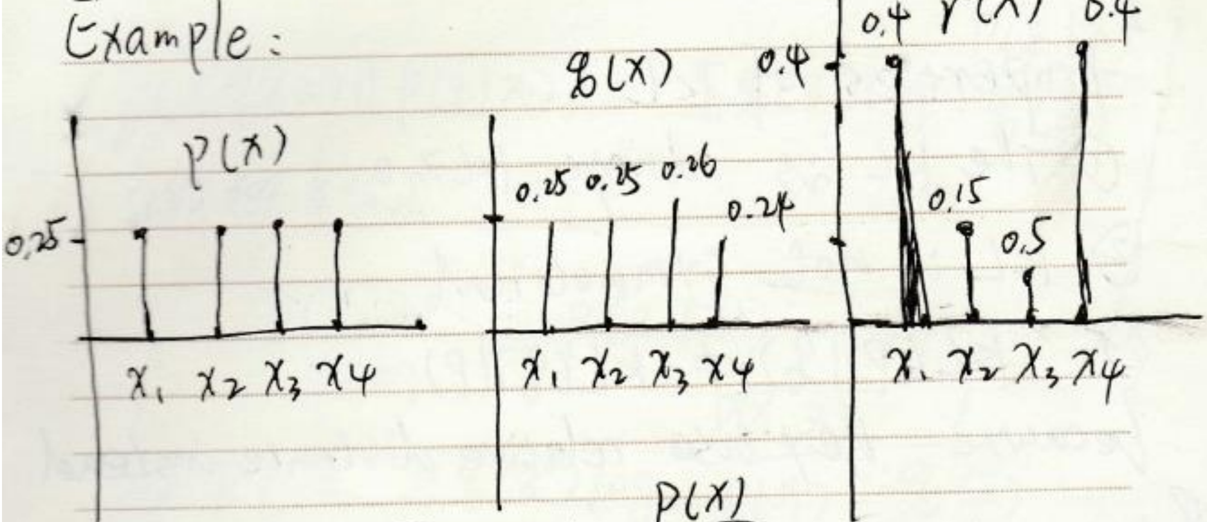
$$KL(P\|Q) = \sum P(x)\log P(x) - \sum P(x)\log Q(x)$$

$$= \sum P(x)\left[\log P(x) - \log Q(x)\right]$$

$$\left.\begin{aligned} &= \sum P(x)\log\frac{P(x)}{Q(x)} \\ &= -\sum P(x)\log\frac{Q(x)}{P(x)} \end{aligned}\right] \text{identical}$$

KL divergence

Example:



$$KL(P\|q) = \sum p(x) \log \frac{p(x)}{q(x)}$$

$$= p(x_1) \log \frac{p(x_1)}{q(x_1)} + p(x_2) \log \frac{p(x_2)}{q(x_2)} + p(x_3) \log \frac{p(x_3)}{q(x_3)}$$

$$+ p(x_4) \log \frac{p(x_4)}{q(x_4)}$$

$$= \cancel{p(x)}\ 0.25 \log \frac{0.25}{0.25} + 0.25 \log \frac{0.25}{0.25} + 0.25 \log \frac{0.25}{0.26}$$

$$+ 0.25 \log \frac{0.25}{0.24} = \boxed{0.0004} \Rightarrow \text{small}$$

$$KL(P\|r) = \sum p(x) \log \frac{p(x)}{q(x)}$$

$$= p(x_1) \log \frac{p(x_1)}{r(x_1)} + p(x_2) \log \frac{p(x_2)}{r(x_2)} + p(x_3) \log \frac{p(x_3)}{r(x_3)}$$

$$+ p(x_4) \log \frac{p(x_4)}{r(x_4)}$$

$$= 0.25 \log \frac{0.25}{0.4} + 0.25 \log \frac{0.25}{0.15} + 0.25 \log \frac{0.25}{0.5}$$

$$+ 0.25 \log \frac{0.25}{0.4} = \boxed{0.295} \Rightarrow \text{large}$$

$$\Rightarrow KL(P\|q) \lll KL(P\|r)$$

2.

Properties of KL

① The KL is always $KL \geq 0$

② KL is not symmetrical

or $KL(P\|q) \neq KL(q\|P)$

because they use relative distance instead

※

3. Relationship between $\ln p(x)$ and KL

- Let's say we have a distribution
  $P(z|x)$, We don't know this

- So we use $q(z)$ to estimate $P(z|x)$

- When using estimation, we can use KL div to measure the quality of the estimation

$$KL(q(z)\|p(z|x)) = -\sum q(z)\log\frac{p(z|x)}{q(z)}$$

条件概率公式

$$\frac{p(x,z)}{p(x)} = p(z|x)$$

$$\Rightarrow KL = -\sum q(z)\log\frac{p(x,z)}{p(x)}\cdot\frac{1}{q(z)}$$

$$\Rightarrow KL = -\sum q(z)\log\frac{p(x,z)}{q(z)}\cdot\frac{1}{p(x)}$$

$-\ln p(x)^{-1}$

$$\Rightarrow KL = -\sum q(z)\left[\log\frac{p(x,z)}{q(z)} + \log\frac{1}{p(x)}\right]$$

$$\Rightarrow KL = -\sum q(z)\left[\log\frac{p(x,z)}{q(z)} - \log p(x)\right]$$

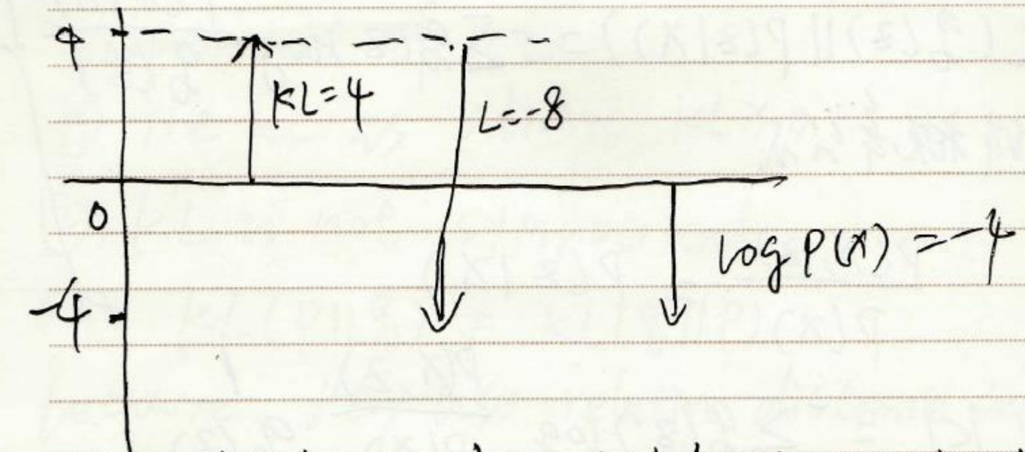
$\cancel{\Rightarrow KL = -\sum q\log}$

$$\Rightarrow KL = -\sum q(z)\log\frac{p(x,z)}{q(z)} + \log p(x)\underbrace{\sum q(z)}_{\text{is}=1}$$

$$\Rightarrow \underset{\text{always } +}{KL} + \underbrace{\sum q(z)\log\frac{p(x,z)}{q(z)}}_{= -KL(q(z)\|p(x,z)) \text{ always } -} = \underbrace{\log p(x)}_{\text{always } -}$$

$\because 0\le p<1 \therefore \log p<0$

$$K + L = \log p(x)$$



L is the lower bound b/c it controls the KL divergence.

∴ By making the lower bound L less negative, We reduce KL divergence.

When reduce KL divergence, the estimation between $q(z)$ and $P(z|x)$ becomes better.

∴ When we are approximating a conditional prob $P(z|x)$ using $q(z)$.

- Instead of minimizing KL Divergence, it's the same as maximizing L.

$$L = \sum q(z) \log \frac{P(x, z)}{q(x)} \uparrow$$

$$KL = -\sum q(z) \log \frac{P(z|x)}{q(z)} \downarrow$$