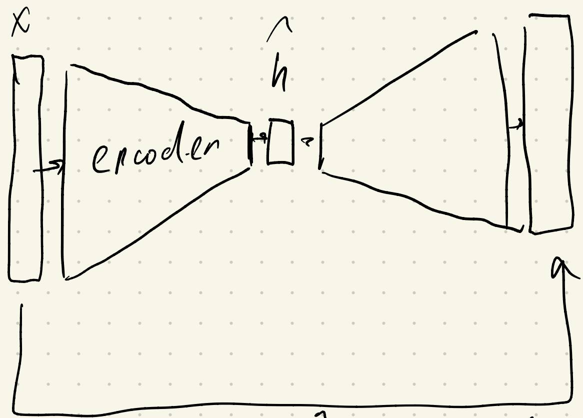
$$-\overset{t_{true}}{P(x)}\log_2(\overset{t_{pred}}{\hat{P}(x)}) - (1-\overset{t_{true}}{P(x)})\log_2(1-\overset{t_{pred}}{\hat{P}(x)})$$

$$P(t|x) = P_t \qquad P_{t(i=k)} = 1 \qquad P_{t(i\neq k)} = 0$$

$$-\log_2(\hat{P}(X)_{(i=k)}) = KL(P, \hat{P})$$

$\hat{x}$



$$h \sim Ber(0.05)$$

$$\|x-\hat{x}\|_2 + KL(P, \hat{P})$$

$$KL(P\|\hat{P}) = \sum_x -P(x)\log_2\left(\frac{\hat{P}(x)}{P(x)}\right)$$

$$x \sim P \qquad I(x) = -\log_2(P(x))$$

$$H(P) = \mathop{\mathbb{E}}_{x\sim P} I(x) = \sum_x -P(x)\log_2(P(x))$$

$$H(P,\hat{P}) = \mathbb{E}_{x\sim P} I_{\hat{P}}(x) = \sum_x -P(x)\log_2 \hat{P}(x) =$$

$$= -\sum_x P(x)\left(\log_2 \hat{P}(x) + \log_2 P(x) - \log_2(P(x))\right)$$

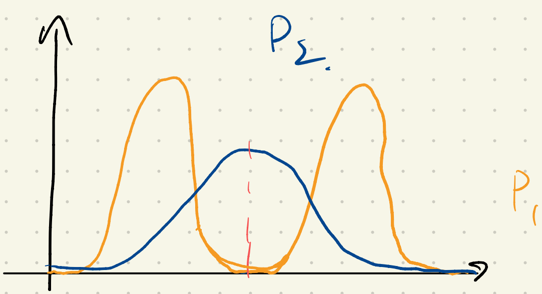
$$= H(P) - \sum\left(\log_2(\hat{P}(x)) - \log_2(P(x))\right)P(x)$$

$$= H(P) - \underbrace{\sum_x P(x)\log_2 \frac{\hat{P}(x)}{P(x)}}_{KL(P\|\hat{P})}$$

1) $\forall P$. $KL(P\|P) = 0$, $KL(P_1\|P_2) = 0$

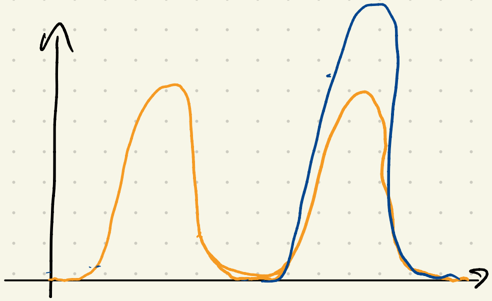
$\rightarrow P_1 \equiv P_2$.

2) $\forall P_1; P_2$ $KL(P_1\|P_2) = KL(P_2\|P_1)$.



$$KL(P_1\|P_2) \rightarrow \min$$

$$-\sum P_1(x)\log\frac{P_2(x)}{P_1(x)}.$$



P_1

$$KL(P_2 \| P_1) \xrightarrow[\min]{} 0$$

$$-\sum P_2(x) \log_2 \frac{P_1(x)}{P_2(x)}$$