Theory Questions

Since,

$$H(P, Q) = H(P) + D_{KL}(P \| Q)$$

$H(P, Q) = D_{KL}(P \| Q)$ can occur when

$H(P) = 0$

→ ~~$-E_{y \sim P}[\log P(y)] = 0$~~

Each term

→ ~~$= E_{y \sim P}[\log P(y)] = 0$~~

→ $-E_{y \sim P}[\log P(y|x)] = 0$

Since $0 \leq P(y|x) \leq 1$, $0 \leq -\log P(y|x) < \infty$

So each of the terms $-y \log P(y|x)$ is ≥ 0.

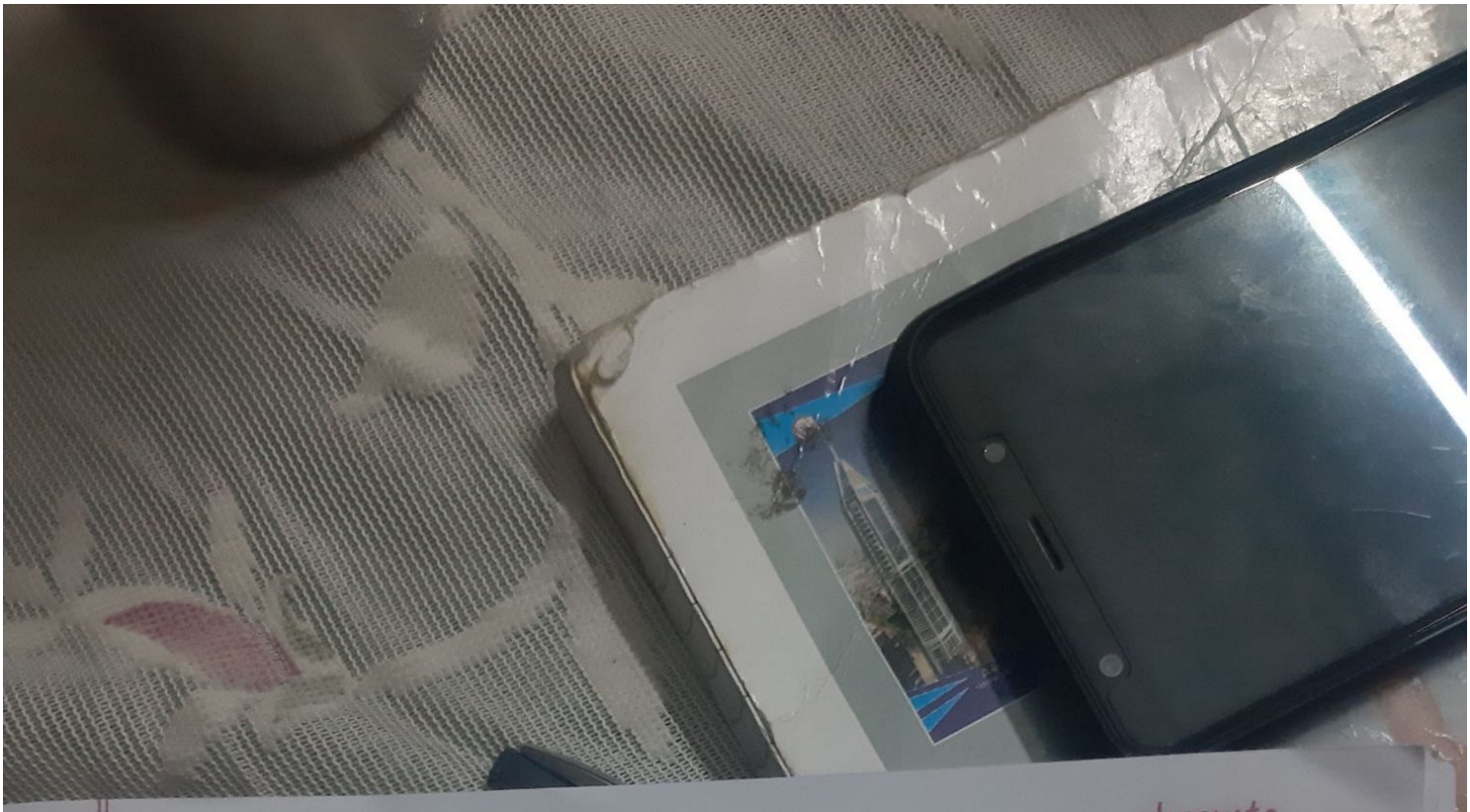
Sum of positive terms is 0 when each of the term is 0.

$$-y \log P(y|x) = 0$$

~~when $y = 1$, $P(y|x) = 1$ or $y = 0$, $P(y|x) = 0$~~.

Assumptions

1. Given the dataset for each [illegible]



If we use 1-hot encoding to represent true values then all the terms $y_i \log P(y_i|x)$ will be 0. provided for the label $y_j = 1$, $P(y_j|x)$ is not predicted as 0 by our model.