ps3-2-solve

May 28, 2021

1 (a)

$$
\begin{aligned}
D_{KL}(P \parallel Q) &= \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} \\
&= \sum_{x \in \mathcal{X}} P(x)(-\log \frac{Q(x)}{P(x)}) \\
&= E[-\log \frac{Q(x)}{P(x)}] \\
&\geq -\log E[\frac{Q(x)}{P(x)}] \\
&= -\log \sum_{x \in \mathcal{X}} P(x) \frac{Q(x)}{P(x)} \\
&= -\log \sum_{x \in \mathcal{X}} Q(x) \\
&= -\log 1 \\
&= 0
\end{aligned}
$$

Beacuse $f(x) = -\log x$ is strictly convex, then

$$
\begin{aligned}
D_{KL}(P \parallel Q) &= 0 \\
E[-\log \frac{Q(x)}{P(x)}] &= -\log E[\frac{Q(x)}{P(x)}]
\end{aligned}
$$

implies that

$$\begin{aligned}
\frac{Q(x)}{P(x)} &= E[\frac{Q(x)}{P(x)}] \\
\frac{Q(x)}{P(x)} &= \sum_{x\in\S} P(x)\frac{Q(x)}{P(x)} \\
\frac{Q(x)}{P(x)} &= \sum_{x\in\S} Q(x) \\
\frac{Q(x)}{P(x)} &= 1 \\
Q(x) &= P(x)
\end{aligned}$$

2 (b)

$$\begin{aligned}
D_{KL}(P(X,Y) \parallel Q(X,Y)) &= \sum_x \sum_y P(x,y) \log \frac{P(x,y)}{Q(x,y)} \\
&= \sum_x \sum_y P(y|x)P(x) \log \frac{P(y|x)P(x)}{Q(y|x)Q(x)} \\
&= \sum_x \sum_y P(y|x)P(x) \log \frac{P(x)}{Q(x)} \\
&\quad + \sum_x \sum_y P(y|x)P(x) \log \frac{P(y|x)}{Q(y|x)} \\
&= \sum_x P(x) \log \frac{P(x)}{Q(x)} \\
&\quad + \sum_x P(x) \big(\sum_y P(y|x) \log \frac{P(y|x)}{Q(y|x)} \big) \\
&= D_{KL}(P(X) \parallel Q(X)) + D_{KL}(P(Y|X) \parallel Q(Y|X))
\end{aligned}$$

3 (c)

$$
\begin{aligned}
\arg\min_\theta D_{KL}(\hat{P} \parallel P_\theta) &= \arg\min_\theta \sum_x \hat{P}(x) \log \frac{\hat{P}(x)}{P_\theta(x)} \\
&= \arg\min_\theta \sum_x \hat{P}(x) \log \hat{P}(x) - \sum_x \hat{P}(x) \log P_\theta(x) \\
&= \arg\min_\theta \big(- \sum_x \hat{P}(x) \log P_\theta(x) \big) \\
&= \arg\max_\theta \sum_x \hat{P}(x) \log P_\theta(x) \\
&= \arg\max_\theta \sum_x \frac{1}{m} \sum_{i=1}^{m} 1\{x^{(i)} = x\} \log P_\theta(x) \\
&= \arg\max_\theta \frac{1}{m} \sum_{i=1}^{m} \sum_x 1\{x^{(i)} = x\} \log P_\theta(x) \\
&= \arg\max_\theta \frac{1}{m} \sum_{i=1}^{m} \log P_\theta(x^{(i)}) \\
&= \arg\max_\theta \sum_{i=1}^{m} \log P_\theta(x^{(i)})
\end{aligned}
$$

[ ]: