4. $f(x) = -\log(x)$ 라 하면

$f''(x) = \frac{1}{x^2}$ $\quad$ $f''(x)$는 $x>0$인 구간에서 모두 0보다 크다.

$\therefore$ $-\log(x)$는 strictly convex

※ $g''(x)>0$ 이면 $g(x)$는 strictly convex 증명

$$\frac{g(\lambda x_1+(1-\lambda)x_2)-g(x_1)}{(1-\lambda)(x_2-x_1)} = g'(x_3) \quad \exists x_3 \in (x_1, \lambda x_1+(1-\lambda)x_2) \text{ by M.V.T}$$

$$\frac{g(x_2)-g(\lambda x_1+(1-\lambda)x_2)}{\lambda(x_2-x_1)} = g'(x_4) \quad \exists x_4 \in (\lambda x_1+(1-\lambda)x_2, x_2) \text{ by M.V.T}$$

$$\frac{g'(x_4)-g'(x_3)}{x_4-x_3} = g''(\alpha) > 0 \quad \exists \alpha \in (x_3, x_4) \text{ by M.V.T}$$

즉, $(1-\lambda)(g(x_2)-g(\lambda x_1+(1-\lambda)x_2)) > \lambda(g(\lambda x_1+(1-\lambda)x_2)-g(x_1))$

$\therefore$ $\lambda g(x_1)+(1-\lambda)g(x_2) > g(\lambda x_1+(1-\lambda)x_2)$

$q_i=0$, $p_i>0$ 이면 $D_{KL}(p\|q)=\infty$ 이므로 $p_i>0$ 이면 $q_i>0$ 이라고 가정. $P(I=i)=p_i$ 라 하면

$$D_{KL}(p\|q) = E_I\left[-\log\left(\frac{q_I}{p_I}\right)\right] \ge -\log\left(E_I\left[\frac{q_I}{p_I}\right]\right)$$
$$= -\log\sum q_i = 0$$

$\therefore$ $D_{KL}(p\|q) \ge 0$

5. $-\log(x)$가 strictly convex하고, $p \neq q$ 이기 때문에

$\frac{q_I}{p_I}$는 contant가 아니다.

$$D_{KL}(p \| q) = E_I\left[-\log \frac{q_I}{p_I}\right] > -\log\left(E_I\left[\frac{q_I}{p_I}\right]\right) = 0$$

$$\therefore D_{KL}(p \| q) > 0$$

6. $f_\theta(x) = u_1 \sigma(a_1 x + b_1) + u_2 \sigma(a_2 x + b_2) + \cdots + u_j \sigma(a_j x + b_j)$

$$\nabla_u f_\theta(x) = \begin{bmatrix} \frac{\partial f_\theta(x)}{\partial u_1} \\ \vdots \\ \frac{\partial f_\theta(x)}{\partial u_j} \end{bmatrix} = \begin{bmatrix} \sigma(a_1 x + b_1) \\ \vdots \\ \sigma(a_j x + b_j) \end{bmatrix} = \sigma(ax + b)$$

$$\nabla_b f_\theta(x) = \begin{bmatrix} \frac{\partial f_\theta(x)}{\partial b_1} \\ \vdots \\ \frac{\partial f_\theta(x)}{\partial b_j} \end{bmatrix} = \begin{bmatrix} u_1 \sigma'(a_1 x + b_1) \\ \vdots \\ u_j \sigma'(a_j x + b_j) \end{bmatrix} = \mathrm{diag}(\sigma'(ax+b))u$$

$$\nabla f_a f_\theta(x) = \begin{bmatrix} \frac{\partial f_\theta(x)}{\partial a_1} \\ \vdots \\ \frac{\partial f_\theta(x)}{\partial a_j} \end{bmatrix} = \begin{bmatrix} u_1 \sigma'(a_1 x + b_1) \cdot x \\ \vdots \\ u_j \sigma'(a_j x + b_j) \cdot x \end{bmatrix} = \mathrm{diag}(\sigma'(ax+b)) \cdot ux$$