4. Put $f : \mathbb{R} \to \mathbb{R}$ It is clear that $\mathbb{R}$ is convex.

$x \mapsto -\log x$ .

Then $f''(x) = \frac{1}{x^2} > 0$ .

By Taylor, we get $\begin{cases} f(x_1) \geq f(\eta x_1 + (1-\eta)x_2) + f'(\eta x_1 + (1-\eta)x_2)\,(1-\eta)\,(x_2 - x_1) \\ f(x_2) \geq f(\eta x_1 + (1-\eta)x_2) + f'(\eta x_1 + (1-\eta)x_2)\,\eta\,(x_1 - x_2) \end{cases}$

$\Rightarrow \quad f(x_1) + (1-\lambda)f(x_2) \geq f(\eta x_1 + (1-\eta)x_2), \quad \eta \in (0,1), \quad \forall x_1, x_2 \in \mathbb{R}$

$\therefore \; f(x) = -\log x$ is convex.

Now, $D_{KL}(p \| q) = \mathbb{E}_I\left(\log \frac{p_I}{q_I}\right), \qquad p(I=i) = p_i$

$= \mathbb{E}_I\left(-\log \frac{q_I}{p_I}\right)$

By Jensen's, $\geq -\log \mathbb{E}_I\left(\frac{q_I}{p_I}\right)$

$= -\log \sum_i \cancel{p_i} \frac{q_i}{\cancel{p_i}}$

$= -\log 1$

$= 0$

$\square$

5. Put $f : \mathbb{R} \to \mathbb{R}$ It is clear that $\mathbb{R}$ is convex.

$$\alpha \mapsto -\log \alpha$$

$\because$ $f''(\alpha) = \frac{1}{\alpha^2} > 0$.

By Taylor, we get
$$\begin{cases} f(\alpha_1) > f(\eta\alpha_1 + (1-\eta)\alpha_2) + f'(\eta\alpha_1 + (1-\eta)\alpha_2)\,(1-\eta)\,(\alpha_1 - \alpha_2) \\ f(\alpha_2) > f(\eta\alpha_1 + (1-\eta)\alpha_2) + f'(\eta\alpha_1 + (1-\eta)\alpha_2)\,\eta\,(\alpha_2 - \alpha_1) \end{cases}$$

$$\Rightarrow \quad f(\alpha_1) + (1-\lambda) f(\alpha_2) > f(\eta\alpha_1 + (1-\eta)\alpha_2), \quad \eta \in (0,1), \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}$$

$\alpha_1 \neq \alpha_2$.

$\therefore$ $f(\alpha) = -\log \alpha$ is strictly convex.

Now,
$$D_{KL}(p \| q) = \mathbb{E}_I\left(\log \frac{p_I}{q_I}\right), \qquad P(I = i) = p_i, \quad p_I \neq q_I$$

$$= \mathbb{E}_I\left(-\log \frac{q_I}{p_I}\right)$$

By Strict Jensen's,
$$> -\log \mathbb{E}_I\left(\frac{q_I}{p_I}\right)$$

$$= -\log \sum_i \cancel{p_i} \frac{q_i}{\cancel{p_i}}$$

$$= -\log 1$$

$$= 0$$

$\square$

6.

i) Show $\nabla_u f_\theta(x) = \sigma(ax+b)$

$$\frac{\partial}{\partial u_i} f_\theta(x) = \frac{\partial}{\partial u_i} \sum_{j=1}^{p} u_j \, \sigma(a_j x + b_j)$$

$$= \frac{\partial}{\partial u_i} \left( u_i \, \sigma(a_i x + b_i) \right)$$

$$= \sigma(a_i x + b_i)$$

$$\therefore \nabla_u f_\theta(x) = \begin{bmatrix} \frac{\partial}{\partial u_1} f_\theta(x) \\ \vdots \\ \frac{\partial}{\partial u_p} f_\theta(x) \end{bmatrix} = \sigma(ax+b)$$

ii) Show $\nabla_b f_\theta(x) = \sigma'(ax+b) \odot u$

$$\frac{\partial}{\partial b_i} f_\theta(x) = \frac{\partial}{\partial b_i} \sum_{j=1}^{n} u_j \, \sigma(a_j x + b_j)$$

$$= \frac{\partial}{\partial b_i} \left( u_i \, \sigma(a_j x + b_i) \right)$$

$$= u_i \, \sigma'(a_i x + b_i) \cdot \frac{\partial}{\partial b_i} (a_j x + b_i)$$

$$= u_i \, \sigma'(a_i x + b_i)$$

$$\therefore \nabla_b f_\theta(x) = \begin{bmatrix} \frac{\partial}{\partial b_1} f_\theta(x) \\ \vdots \\ \frac{\partial}{\partial b_p} f_\theta(x) \end{bmatrix} = \begin{bmatrix} \sigma'(a_1 x + b_1) \cdot u_1 \\ \vdots \\ \sigma'(a_p x + b_p) \cdot u_p \end{bmatrix} = \sigma'(ax+b) \odot u =$$

$$= \begin{bmatrix} \sigma'(a_1 x + b_1) & & 0 \\ & \ddots & \\ 0 & & \sigma'(a_p x + b_p) \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix} = \mathrm{diag}(\sigma'(ax+b))\, u$$

iii) Show $\nabla_a f_\theta(x) = (\sigma'(ax+b) \odot u) x$

$$\frac{\partial}{\partial a_i} f_\theta(x) = \frac{\partial}{\partial a_i} \sum_{j=1}^{p} u_j \, \sigma(a_j x + b_j) = \frac{\partial}{\partial a_i} \left( u_i \, \sigma(a_i x + b_i) \right)$$

$$= u_i \, \sigma'(a_i x + b_i) \frac{\partial}{\partial a_i} (a_i x + b_i) = u_i \, \sigma'(a_i x + b_i)\, x$$

$$\nabla_a f_\theta(x) = \begin{bmatrix} \frac{\partial}{\partial a_1} f_\theta(x) \\ \vdots \\ \frac{\partial}{\partial a_p} f_\theta(x) \end{bmatrix} = \begin{bmatrix} \sigma'(a_1 x + b_1) \cdot u_1 \cdot x \\ \vdots \\ \sigma'(a_p x + b_p) \cdot u_p \cdot x \end{bmatrix} = (\sigma'(ax+b) \odot u) x$$

$$= \begin{bmatrix} \sigma'(a_1 x + b_1) & & 0 \\ & \ddots & \\ 0 & & \sigma'(a_p x + b_p) \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix} x = \mathrm{diag}(\sigma'(ax+b))\, u\, x$$

$\square$