4.1-b. We wanna show $p(y_w \succ y_l \mid x) = \sigma\left(\beta \log \frac{\pi_\phi^*(y_w|x)}{\pi_{ref}(y_w|x)} - \beta \log \frac{\pi_\phi^*(y_l|x)}{\pi_{ref}(y_l|x)}\right)$

We have $p(y_w \succ y_l \mid x) = \dfrac{\exp r^*(x, y_w)}{\exp r^*(x, y_w) + \exp r^*(x, y_l)}$ — ①

$\pi^*(y|x) = \frac{1}{Z} \pi_{ref}(y|x) \exp\left(\frac{1}{\beta} r^*(x,y)\right)$ → Putting log & solving

$r^*(x,y) = \beta \log \frac{\pi^*(y|x)}{\pi_{ref}(y|x)} + \beta \log Z$ — ②

Putting ② in ①

$$p(y_1 \succ y_2 \mid x) = \frac{\exp\left(\beta \log \frac{\pi^*(y_w|x)}{\pi_{ref}(y_w|x)} + \beta \log Z\right) \rightarrow ⓐ}{\exp\left(\beta \log \frac{\pi^*(y_w|x)}{\pi_{ref}(y_w|x)} + \beta \log Z\right) + \exp\left(\beta \log \frac{\pi^*(y_l|x)}{\pi_{ref}(y_l|x)} + \beta \log Z\right)}$$

Dividing ⓐ in num & denom

$$= \frac{1}{1 + \exp\left(\beta \log \frac{\pi^*(y_l|x)}{\pi_{ref}(y_l|x)} - \beta \log \frac{\pi^*(y_w|x)}{\pi_{ref}(y_w|x)}\right)} \quad — ⓑ$$

The sigmoid function is $\sigma(z) = \frac{1}{1+\exp(-z)}$ so, we can write ⓑ in this form.

Hence,

$$\boxed{p(y_w \succ y_l \mid x) = \sigma\left(\beta \log \frac{\pi_\phi^*(y_w|x)}{\pi_{ref}(y_w|x)} - \beta \log \frac{\pi_\phi^*(y_l|x)}{\pi_{ref}(y_l|x)}\right)}$$

Hence, proved.