# Problem 4

Consider a classification problem with:

- A target variable $Y$ with $K$ classes.
- A binary feature $X \in \{0, 1\}$ used for splitting.

We define:

- $H(Y)$: Entropy of the target variable $Y$.
- $H(Y|X)$: Conditional entropy of $Y$ given $X$.
- $I(X;Y)$: Mutual information between $X$ and $Y$, which represents the **information gain** from splitting on $X$.

The mutual information is defined as:

$$I(X;Y) = H(Y) - H(Y|X)$$

From information theory, mutual information is symmetric:

$$I(X;Y) = I(Y;X) = H(X) - H(X|Y)$$

Since $X$ is binary, we know:

$$H(X) \leq 1$$

and the conditional entropy $H(X|Y)$ is non-negative:

$$H(X|Y) \geq 0$$

Therefore:

$$I(X;Y) = H(X) - H(X|Y) \leq H(X) \leq 1$$

Thus:

$$H(Y) - H(Y|X) = I(X;Y) \leq 1$$