

### Exercise 1: Entropy in Binary Classification

Suppose we are facing a binary classification learning problem with the feature space  $\mathcal{X} = \{0, 1\}$  and label space  $\mathcal{Y} = \{-1, 1\}$ . Let us assume that the relationship between the features and labels is specified by

$$y = 2 \max\{x, \epsilon\} - 1$$

where  $\mathbb{P}_x = \text{Ber}(\theta_x)$ ,  $\mathbb{P}_\epsilon = \text{Ber}(\theta_\epsilon)$ , and  $x$  and  $\epsilon$  are independent.

(a) What is  $H(\epsilon)$ ?

(b) What is  $H(y|x)$ ?

(c) What is  $H(y, x)$ ?

(d) Now if  $\epsilon = \psi(x)$  for some deterministic function  $\psi : \{0, 1\} \rightarrow \{0, 1\}$ . What are the answers to the previous questions in this case?