1. The distribution of (X, Y) is given by

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 0$	0.2	0.2	0.1
$X = 1$	0.5	0	0

   (a) Find the entropies (X), (Y) and the joint entropy (X, Y).
   (b) Find conditional entropy $(Y \mid X)$.
   (c) What is a maximal value of conditional entropy $(Y \mid X)$, if X has two possible values and Y — three possible values?

2. Consider a toy dataset $y = (10, 10, 30, 30, 30, 40)$ and $x = (1, 2, 3, 4, 5, NA)$.
   Build the regression true based on the following rules:
   - The tree has two splitting nodes.
   - Each node is split to minimize sum of squared residuals.
   - If there are many possible nodes to split choose the node with maximal sum of squared residuals drop.
   - For missing observations (NA) consider both split directions.
   - In case of many splits with the same sum of squared residuals drop choose the split with minimal threshold on x.

3. Consider the simple logit model for binary $y_i \in \{0, 1\}$. We assume that $(y_i \mid x)$ are independent and

$$\mathbb{P}(y_i = 1 \mid x) = \frac{\exp(\beta_1 + \beta_x x_i)}{1 + \exp(\beta_1 + \beta_x x_i)}.$$

   You don't know the true values of β_1 and β_x and you would like to estimate them using maximum likelihood.
   (a) Write the likelihood function in the general form.
   (b) Write the first order conditions $\partial \ln L / \partial \hat{\beta}_1 = 0$ and $\partial \ln L / \partial \hat{\beta}_x = 0$.
   (c) You have 200 observations and 150 of them are $y_i = 1$.
   What will be the value of a sum $\sum_{i=1}^{200} \hat{\mathbb{P}}(y_i = 1 \mid x)$?

   Hint: the formula for $\hat{\mathbb{P}}(\ldots)$ is similar to $\mathbb{P}(\ldots)$, just β-s are replaced by $\hat{\beta}$-s. You can answer (c) without solving the system of first order conditions.

4. Random variables y_1, y_2, ..., y_n is a random sample from a continuous distribution. Let y_1^*, y_2^*, ..., y_n^* be the first bootstrap sample.
   (a) Find $\mathbb{P}(y_1 = y_2^*)$, $\mathbb{P}(y_1 = y_2)$, $\mathbb{P}(y_1^* = y_2^*)$.
   (b) Find $\mathbb{P}($ all y_i^* are equal$)$.
   (c) Find $\mathbb{P}(\max\{y_1, y_2, \ldots, y_n\} = \max\{y_1^*, y_2^*, \ldots, y_n^*\})$.
   (d) What is the probability that all original observations are present in the bootstrap sample?