

Фамилия:.....

Имя:.....

Группа:.....

Задача №1

У вас имеется выборка из $n = 8$ наблюдений, характеризующих успешность стартапов в зависимости от наличия рекламной кампании и опытных членов команды, а также того, связан ли стартап с тематикой здоровья.

Успех	Реклама	Опыт	Здоровье
1	1	1	0
0	1	1	1
1	0	1	1
0	0	0	1
1	1	1	0
0	0	0	0
1	0	1	0
0	0	1	0

Вы обучаете решающее дерево глубины 2, прогнозирующее успех стартапа с помощью всех имеющихся в данных признаков (реклама, опыт и здоровье). В качестве критерия разбиения используется энтропия. Прогнозируется, что стартап окажется успешным, если условная вероятность этого события превышает 0.4.

Подсказка: вместо логарифма с основанием 2 эквивалентно использовать натуральный логарифм, необходимые значения которого указаны ниже:

$$\begin{aligned}
 \ln(1/8) &\approx -2.079 & \ln(2/8) &\approx -1.386 & \ln(3/8) &\approx -0.981 & \ln(4/8) &\approx -0.693 \\
 \ln(5/8) &\approx -0.470 & \ln(6/8) &\approx -0.288 & \ln(7/8) &\approx -0.134 & \ln(8/8) &= -0.000 \\
 \ln(1/6) &\approx -1.792 & \ln(2/6) &\approx -1.099 & \ln(4/6) &\approx -0.405 & \ln(5/6) &\approx -0.182 \\
 \ln(1/5) &\approx -1.609 & \ln(2/5) &\approx -0.916 & \ln(3/5) &\approx -0.511 & \ln(4/5) &= -0.223
 \end{aligned}$$

1. Изобразите обученное решающее дерево графически, в каждом листе указав долю успешных стартапов. Обязательно приведите в решении подробные расчеты, с помощью которых было построено дерево. **(20 баллов)**
2. С использованием построенного решающего дерева для обучающей выборки запишите матрицу ошибок (confusion matrix) и рассчитайте F1-метрику. **(15 баллов)**
3. Используя обученное ранее решающее дерево с помощью S-learner оцените средний эффект воздействия тематики здоровья на успех стартапа. **(20 баллов)**
4. Предположим, что выполняется предпосылка об условной независимости, а также, что потенциальные исходы успеха (в зависимости от здоровья), Успех_{0i} и Успех_{1i} , независимы между собой при любых фиксированных значениях признаков. Используя обученное решающее дерево получите состоятельную оценку математического ожидания $E\left(\frac{\text{Успех}_{1i}}{1+\text{Успех}_{0i}}\right)$. **(20 баллов)**

Задача №2

Имеется нейросеть, включающая всего $n = 1$ наблюдение по 2 признакам: $x_1 = 1$ (первый признак) и $x_2 = 2$ (второй признак). Значение целевой переменной равняется $y = 145$. Имеется лишь 1 скрытый слой с 2 нейронами. В качестве функции активации в скрытом и выходном слоях используется ReLU. Применяется квадратичная функция потерь. В нейросети нет смещений (констант) и все ее параметры равняются 5.

1. Изобразите графически описанную нейросеть. (2 балла)
2. Рассчитайте значение функции потерь данной нейросети при заданных значениях весов. (8 баллов)
3. Для обучения нейронной сети используется градиентный спуск. После одной его итерации вес, с которым второй признак входит в первый нейрон первого скрытого слоя, оказался равен 0. Найдите скорость обучения градиентного спуска. (15 баллов)