ФИО, курс: ____.

Правила

Строго запрещено:

- переговариваться (с любой целью),
- пользоваться устройствами связи (с любой целью — например, в качестве калькулятора).
- списывать.

Нарушение любого из этих пунктов влечет удаление с контрольной работы.

Пожалуйста, пишите подробные решения!

Желаем удачи!

Задача 1. Рассмотрим задачу двухклассовой классификации с двумя признаками. Обучающая выборка представлена на рис. 1, классы обозначены крестиком и ноликом.

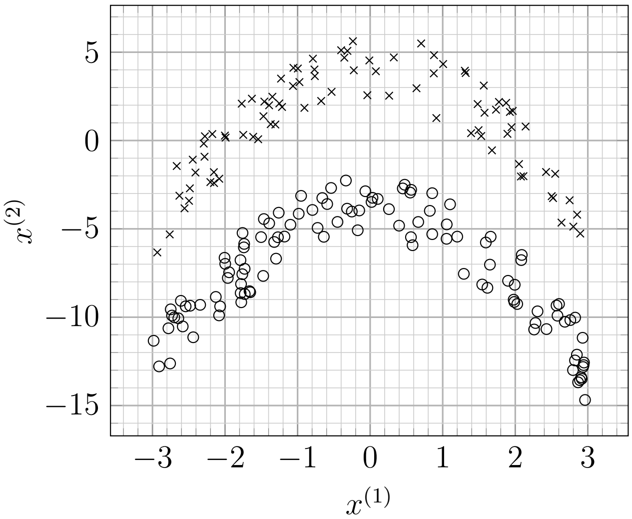

Рис. 1: Рисунок к задаче 1

a. Придумать новый признак $x^{(3)}$, значение которого вычисляется по $x^{(1)}$ и $x^{(2)}$, такой, что для нового признакового пространства $(x^{(1)}, x^{(2)}, x^{(3)})$ можно найти такие веса $w = (w_1, w_2, w_3)$ в логистической регрессии
$$\log \frac{p}{1-p} = \langle w, x \rangle,$$
что accuracy на обучающей выборке будет равно 100%. Здесь p — это вероятность, что объект относится к классу, обозначаемому крестиком. Если $p > 1/2$, предсказывается крестик, иначе нолик.

b. Найти эти веса.

c. Закрасить на рисунке 1 область, точки в которой построенный таким образом классификатор отнесёт к крестикам.

Задача 2. Снова рассмотрим задачу двухклассовой классификации с двумя признаками. Обучающая выборка представлена на рис. 2, классы обозначаются заполненным и пустым кружочком.

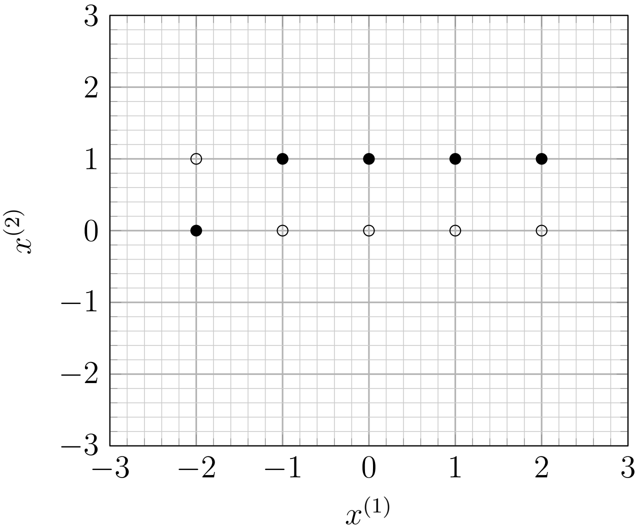

Рис. 2: Рисунок к задаче 2.

Компьютер строит решающее дерево с помощью обычного «жадного» алгоритма. На каждом этапе объекты делятся на два подмножества в зависимости от результата сравнения какого-то из признаков $x_i^{(j)}$, $j = 1, 2$ с порогом t. Значения j и t выбираются таким образом, чтобы минимизировать функцию

$$Q(L, R) = |L| \cdot I_G(L) + |R| \cdot I_G(R),$$

где $I_G(M)$ — значение Gini impurity function на соответствующем множестве, L и R — части, на которые делится выборка, $|R|$ и $|L|$ — количество элементов в соответствующем множестве. Значения t выбираются ровно посередине между ближайшими значениями $x_i^{(j)}$ в выборке. Если есть несколько одинаково оптимальных с точки зрения функции Q значений t, выбирается меньшее из них. Дерево строится до тех пор, пока в каждом листе все объекты не окажутся принадлежащими одному классу. Какое дерево получится у компьютера? Обоснуйте выбор признака j, по которому происходит разделение, и значение порога t на каждом шаге.

Задача 3. Рассмотрим задачу регрессии с одним признаком. Обучающая выборка состоит из трёх объектов и приведена в таблице

i	x_i	y_i
1	1	-4
2	2	3
3	3	-4

Задача решается с помощью случайного леса, состоящего из трёх решающих пней (деревьев глубины 1). Первый пень обучается по выборке, состоящей из наблюдения 1, снова 1 и 2. Второй — по наблюдениям 2, 3 и снова 3, третий — по наблюдениям 1, 3 и снова 3. Предсказанием пня является среднее значение y_i по всем объектам, попавшим в соответствующий лист. Пороговое значение t выбирается путём минимизации суммы квадратов отклонений предсказания от истинного значения (RSS). Значение t выбирается ровно посередине между двумя соседними значениями x_i. Предсказанием всего леса является среднее арифметическое предсказаний каждого из пней.

a. Найдите пороговые значения для каждого из пней.
b. Постройте график зависимости предсказания леса от x.

Задача 4. Задача регрессии решается с помощью градиентного бустинга. На каждом шаге строится алгоритм $a_{n+1}(x) = a_n(x) + b_n(x)$ путём обучения алгоритма b_n. На некотором шаге бустинга построенный к тому моменту алгоритм a_n выдал предсказания, приведенные в таблице в столбце $\hat{y}_i$. Истинные ответы приведены в таблице в столбце y_i.

y_i	$\hat{y}_i$
4	3
-2	-7
-4	-4

В качестве функции потерь используется Huber loss:

$$L(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & |y - \hat{y}| \leq \delta, \\ \delta|y - \hat{y}| - \frac{1}{2}\delta^2, & |y - \hat{y}| > \delta, \end{cases}$$

в которой положили $\delta = 2$. Найти градиент, на который должен обучаться алгоритм b_n.

Задача 5. Рассмотрим многомерный перспептрон с одним скрытым слоем. На скрытом слое в качестве функции активации используется

$$\mathrm{ReLU}(x) = \max\{0, x\}.$$

Про функцию потерь ничего неизвестно, кроме того, что она дифференцируема всюду.

Рассмотрим некоторый нейрон на скрытом слое. Он связан с тремя входными нейронами, выходы которых равны значениям признаков $x^{(1)}, x^{(2)}, x^{(3)}$. Веса связей равны w_1, w_2, w_3. Значения признаков и веса в некоторый момент приведены в таблице.

i	$x^{(i)}$	w_i
1	-3	-1
2	-5	3
3	1	0

Найти производную функции потерь по w_1 в точке, соответствующей указанному значению признаков и весов.

Задача 6. Для анализа изображений используется свёрточная нейронная сеть. Будем считать входное изображение матрицей (x^0_{ij}), $i = 1, \dots, n$, $j = 1, \dots, m$, $x^0_{ij} \in [0, 1]$, каждый элемент задаёт один пиксель, изображение чёрно-белое. Исходное изображение поступает на вход первому свёрточному слою, а выход первого свёрточного слоя сразу поступает на вход второму свёрточному слою (между свёрточными слоями нет никакого max pooling или других преобразований). Значения элементов первого и второго свёрточного слоя обозначаются через x^1_{ij} и x^2_{ij} соответственно, $i = 1, \dots, m$, $j = 1, \dots, m$ (размеры свёрточных слоёв совпадают с размерами исходного изображения). Каждый свёрточный слой использует ядро размером 3×3. Значение элемента с индексом (i, j) на каждом свёрточном слое является результатом примеения свёртки к «окошку» от $(i-1, j-1)$ до $(i+1, j+1)$ (включительно):

$$x^{n+1}_{ij} = \sum_{k=1}^{3} \sum_{l=1}^{3} x^n_{i-2+k, j-2+l} w^n_{kl},$$

где w^n_{kl} — соответствующее свёрточное ядро. Если в формуле возникают значения элементов с индексами меньше 1 или больше m, такие значения считаются равными нулю.

Сколько пикселей исходного изображения влияет на значение элемента x^2_{12} второго свёрточного слоя, если предположить, что все ядра не имеют нулевых весов? (Это множество пикселей называется рецептивным полем (receptive field) данного элемента.)

Задача 7. Рассмотрим задачу кластеризации. Пространство признаков двумерное. Имеется четыре объекта, обозначенные на рис. 3 кружочками. Мы применяем алгоритм K-means и хотим выделить два кластера. Квадратиком и треугольником на картинке обозначены центроиды в начальный момент. Указать, какие объекты к каким кластерам будут относиться и где будут располагаться центроиды после того, как алгоритм K-means сойдётся.

Замечание. После того, как на очередном шаге определено, к какому кластеру какой объект относится, каждый центроид перемещается в такую точку, что сумма квадратов расстояний от неё до всех объектов, относящихся к данному кластеру, минимальна. Метрика стандартная евклидова.

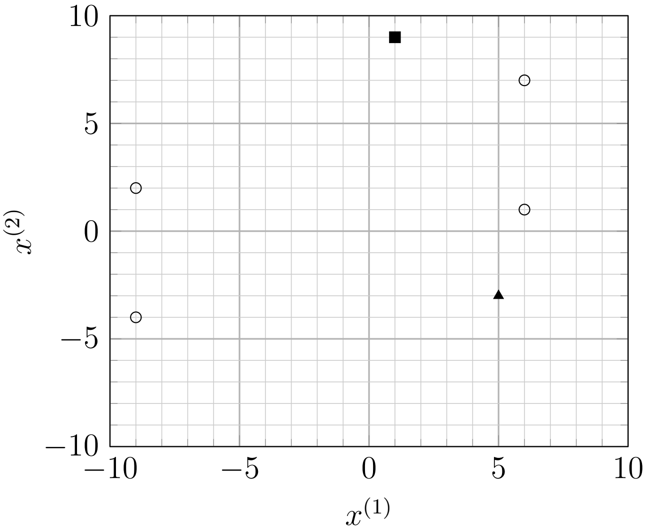

Рис. 3: Рисунок к задаче 7.