ВАЖНО! ДАННЫЙ ВАРИАНТ ЯВЛЯЕТСЯ ТОЛЬКО ДЕМОВЕРСИЕЙ.
НА САМОМ ОТБОРЕ МОГУТ БЫТЬ ДРУГИЕ ЗАДАЧИ И НОВЫЕ ТЕМЫ.
ОСТАЛЬНЫЕ ПРИМЕРЫ ЗАДАНИЙ БУДУТ ДОСТУПНЫ ПОСЛЕ РЕГИСТРАЦИИ — РЕГИСТРИРУЙТЕСЬ И СЛЕДИТЕ ЗА ОБНОВЛЕНИЯМИ.

Задание 1

Рассмотрим простой перцептрон с двумя признаками x_1 и x_2. Он вычисляет значение

$$z = 2x_1 - x_2 - 1.$$

Если $z > 0$, модель предсказывает класс 1, иначе — класс 0.

Выберите все верные утверждения.

(a) Если увеличить x_1 на 1, не меняя x_2, значение z увеличится на 2.

(b) Если одновременно увеличить x_1 на 1, а x_2 на 2, предсказание модели не изменится.

(c) Все точки, удовлетворяющие равенству

$$2x_1 - x_2 - 1 = 0,$$

лежат на границе между двумя классами.

(d) Если одновременно умножить все коэффициенты и свободный член на одно и то же отрицательное число, предсказания модели не изменятся.

Задание 2

Алгоритм **kNN (k-nearest neighbors)** используется для предсказания по похожим объектам.

Для нового объекта алгоритм:

1. вычисляет расстояние от него до объектов обучающей выборки;
2. выбирает k ближайших объектов (правило выбора расстояния можно менять; стандартный вариант – евклидово расстояние);
3. в задаче классификации предсказывает наиболее частый класс среди выбранных соседей.

Выберите все верные утверждения о kNN.

(a) Уменьшение k уменьшает влияние отдельных случайных точек и шума на предсказание модели.

(b) Способ, с помощью которого определяется расстояние между объектами, не влияет на итоговое предсказание kNN.

(c) Добавление случайного признака, не связанного с целевой переменной, может ухудшить качество kNN.

(d) При использовании евклидова расстояния, если один из признаков умножить на очень большое число, он может начать сильнее влиять на выбор ближайших соседей.

Задание 3

Рассмотрим модель линейной регрессии

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b.$$

Качество модели на обучающей выборке оценивается с помощью средней квадратичной ошибки (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$

где y_i – истинное значение для объекта i, а $\hat{y}_i$ – предсказание модели.

Параметры $w_1, \ldots, w_d, b$ выбираются так, чтобы минимизировать MSE на обучающей выборке.

Выберите все верные утверждения.

(a) Если добавить к данным ещё один признак, минимально возможная MSE на обучающей выборке не может увеличиться.

(b) Если все значения одного признака умножить на ненулевую константу и после этого заново обучить модель, минимально возможная MSE обязательно изменится.

(c) Если два признака имеют одинаковые значения для всех объектов, могут существовать разные наборы весов, дающие одинаковые предсказания.

(d) Предсказания линейной регрессии всегда лежат между минимальным и максимальным значениями y в обучающей выборке.

Задание 4

Пусть модель для любого объекта предсказывает одно и то же число c.

Качество модели оценивается с помощью MSE:

$$\text{MSE}(c) = \frac{1}{n} \sum_{i=1}^{n} (y_i - c)^2,$$

где $y_1, \ldots, y_n$ – истинные значения.

1. Для набора значений

   $$2,\ 4,\ 7,\ 11$$

   найдите такое c, при котором MSE минимально.
2. Теперь рассмотрим произвольные значения $y_1, \ldots, y_n$. Найдите значение c, при котором MSE минимально. Обоснуйте ответ.
3. Пусть одно из истинных значений увеличили на Δ, а остальные оставили без изменений. На сколько изменится оптимальное значение c?

Задание 5

Рассмотрим модель **Decision Tree**. Пусть у нас есть N различных точек, для каждой из которых нужно предсказать некоторое значение.

a)

Какой глубины дерева достаточно в худшем случае, чтобы для любых N различных точек можно было построить дерево, в котором каждая точка окажется в отдельном листе?

Почему такое дерево обычно является плохой идеей? Какую проблему это может вызвать?

b)

Рассмотрим один из признаков F. Перед обучением дерева ко всем значениям этого признака прибавили одну и ту же константу t:

$$F' = F + t.$$

После этого дерево обучили заново.

Изменится ли из-за такого преобразования лучшее достижимое качество дерева? Объясните свой ответ.

c)

Пусть максимальная глубина дерева равна 2.

При построении дерева каждое разбиение выбирается жадно: на каждом шаге выбирается разбиение, которое сильнее всего улучшает выбранный критерий качества на текущем шаге.

Гарантирует ли такой подход, что в итоге получится лучшее возможное дерево глубины 2?

Если нет, приведите простой пример, когда выбор лучшего разбиения на первом шаге приводит к худшему итоговому дереву, чем другой выбор первого разбиения. Можно использовать любой стандартный критерий качества разбиения; укажите, какой именно критерий вы используете.

Задание 6

Рассмотрим модель **логистической регрессии** для бинарной классификации. Для объекта с признаками $x_1, \ldots, x_d$ она сначала вычисляет

$$z = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b,$$

а затем преобразует это значение в вероятность класса 1 с помощью функции

$$p = \sigma(z) = \frac{1}{1 + e^{-z}}.$$

Если $p > 0.5$, модель предсказывает класс 1, иначе — класс 0.

a)

Докажите, что

$$p > 0.5$$

тогда и только тогда, когда

$$z > 0.$$

b)

Докажите, что для любого z

$$\sigma(-z) = 1 - \sigma(z).$$

Как связаны вероятности класса 1 для двух объектов, для которых значения до применения функции σ равны z и $-z$?

c)

Могут ли существовать две модели логистической регрессии, которые для **любого** объекта предсказывают один и тот же класс, но при этом не для всех объектов выдают одинаковую вероятность класса 1?

Если да, приведите пример двух таких моделей и докажите, что они обладают этим свойством.

Задание 7

В свёрточных нейронных сетях изображение можно обрабатывать с помощью небольшого набора чисел – **фильтра**.

В этой задаче будем рассматривать одноканальное изображение и фильтр размера 3×3. Фильтр последовательно помещается во все возможные положения внутри изображения с шагом в один пиксель по горизонтали и вертикали. За границы изображения фильтр не выходит.

В каждом положении рассматривается участок изображения размера 3×3. Чтобы получить одно число на выходе, каждый элемент этого участка умножается на соответствующий элемент фильтра, после чего все полученные произведения складываются.

Если участок изображения обозначить через A, а фильтр через W, то результат равен

$$z = \sum_{i=1}^{3} \sum_{j=1}^{3} A_{ij} W_{ij}.$$

Для каждого положения фильтра получается одно число. Все эти числа образуют новую матрицу.

Рассмотрим изображение

$$X = \begin{pmatrix} 1 & 1 & 1 & 4 & 4 \\ 1 & 1 & 1 & 4 & 4 \\ 1 & 1 & 1 & 4 & 4 \\ 2 & 2 & 2 & 5 & 5 \\ 2 & 2 & 2 & 5 & 5 \end{pmatrix}$$

и фильтр

$$W = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Во всех положениях используется один и тот же фильтр W.

a)

Какого размера будет полученная после применения фильтра матрица?

Вычислите все элементы её первой строки.

b)

К каждому элементу исходного изображения прибавили одно и то же число c:

$$X'_{ij} = X_{ij} + c.$$

Изменятся ли значения, полученные после применения фильтра W?

Докажите свой ответ.

c)

Пусть в другой части изображения встретился участок размера 3×3, полностью совпадающий с участком, который уже встречался раньше.

Будет ли результат применения фильтра к этим двум участкам одинаковым?

Объясните почему.

Задание 8

SVM (Support Vector Machine) – алгоритм классификации, который пытается провести границу между объектами разных классов так, чтобы расстояние от этой границы до ближайших обучающих объектов было как можно больше.

Иногда классы нельзя удобно разделить в исходном пространстве признаков. Тогда SVM может рассматривать объекты в некотором новом пространстве. При этом координаты объектов в новом пространстве необязательно вычислять явно: достаточно уметь считать их скалярные произведения с помощью функции **kernel**:

$$k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle.$$

Для объектов $x_1, \ldots, x_n$ можно составить kernel-матрицу

$$K_{ij} = k(x_i, x_j).$$

Корректная kernel-матрица должна быть:

- симметричной:
$$K_{ij} = K_{ji};$$
- положительно полуопределённой (PSD), то есть для любого вектора v
$$v^T K v \geq 0.$$

Рассмотрим kernel-матрицу

$$K = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Известно, что эта матрица является симметричной и PSD.

a)

Докажите, что квадрат расстояния между двумя объектами в новом пространстве можно найти по формуле

$$\|\varphi(x_i) - \varphi(x_j)\|^2 = K_{ii} + K_{jj} - 2K_{ij}.$$

Используя эту формулу, найдите расстояния между всеми парами объектов x_1, x_2, x_3.

Какие объекты находятся ближе всего друг к другу?

b)

Для объекта x SVM вычисляет

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b.$$

Если $f(x) > 0$, модель предсказывает класс $+1$, а если $f(x) < 0$ – класс -1.

Пусть

$$y = (+1, +1, -1),$$

$$\alpha = \left(\frac{1}{2}, \frac{1}{2}, 1\right), \qquad b = 0.$$

Используя матрицу K, найдите значения

$$f(x_1), \qquad f(x_2), \qquad f(x_3)$$

и определите предсказанный класс для каждого объекта.

c)

Докажите, что у любой PSD-матрицы все диагональные элементы неотрицательны:

$$K_{ii} \geq 0.$$

Подсказка: рассмотрите вектор v, у которого все элементы равны нулю, кроме i-го элемента, равного 1.

Как это свойство связано с равенством

$$K_{ii} = \langle \varphi(x_i), \varphi(x_i) \rangle?$$

d)

Из исходной kernel-матрицы построили новую:

$$K' = 3K + 2I,$$

где I – единичная матрица.

Докажите, что K' также является корректной kernel-матрицей. Для этого покажите, что:

1. K' симметрична;
2. для любого вектора v

$$v^T K' v \geq 0.$$

**ВАЖНО! ДАННЫЙ ВАРИАНТ ЯВЛЯЕТСЯ ТОЛЬКО ДЕМОВЕРСИЕЙ.
НА САМОМ ОТБОРЕ МОГУТ БЫТЬ ДРУГИЕ ЗАДАЧИ И НОВЫЕ ТЕМЫ.
ОСТАЛЬНЫЕ ПРИМЕРЫ ЗАДАНИЙ БУДУТ ДОСТУПНЫ ПОСЛЕ РЕГИСТРАЦИИ
— РЕГИСТРИРУЙТЕСЬ И СЛЕДИТЕ ЗА ОБНОВЛЕНИЯМИ.**