ФИО, курс, группа: ____.

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 классификации с двумя классами: A и B. Пространство признаков двумерное. Данные имеют следующий вид: объекты $(-4, 0)$, $(0, 4)$, $(-4, 4)$ имеют класс A и объект $(-2, 2)$ имеет класс B. Будем использовать метод k ближайших соседей (k-NN).

a. (6 баллов) Пусть для точки $(x^{(1)}, x^{(2)})$ k-NN, обученный по всем данным, даёт предсказание $f_k(x^{(1)}, x^{(2)})$. Разбить плоскость на области $\{(x^{(1)}, x^{(2)}) \mid f_k(x^{(1)}, x^{(2)}) = A\}$ и $\{(x^{(1)}, x^{(2)}) \mid f_k(x^{(1)}, x^{(2)}) = B\}$ для $k = 1$ и $k = 3$.

b. (8 баллов) Выбрать оптимальный $k \in \{1, 3\}$ с помощью 4-фолдовой кросс-валидации, метрика — accuracy.

Задача 2. (10 баллов) Рассмотрим задачу линейной регрессии $y_i = w_0 + w_1 x_i + \varepsilon_i$, $x_i \in \mathbb{R}$. Данные имеют следующий вид:

x	y
-3	-4
2	1
2	7
2	4
-3	2

Найти оценку для w_0 и w_1 с помощью метода наименьших квадратов.

Задача 3. (10 баллов) Рассмотрим задачу регрессии с одномерным пространством признаков. Пусть истинный закон генерирования данных описывается следующим образом: все x_i фиксированы и заданы так: $x_i = i - 3$, $i = 1, \dots, 5$, а y_i являются случайными величинами:

$$y_i = |4x_i| + \varepsilon_i,$$

где все ε_i независимы, $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{D}[\varepsilon_i] = 1$. Пусть $f_k(x)$ — предсказание метода k ближайших соседей в точке x. Найти ожидаемую квадратичную ошибку предсказания в точке $x = 0$ для $k = 3$. Представить её в виде суммы шума, смещения и разброса.

Задача 4. (15 баллов) Рассмотрим задачу регрессии с двумерным пространством признаков. Данные генерируются следующим образом: x-составляющие фиксированы и равны $(8, 0)$, $(-8, 0)$, $(0, 1)$, $(0, -1)$, ответы y_i генерируются так:

$$y_i = x_i^{(1)} + 10x_i^{(2)} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

все ε_i независимы. Рассмотрим четыре предсказательные модели. Модель A:

$$\hat{y}_i = \hat{w}_1^A x_i^{(1)} + \hat{w}_2^A x_i^{(2)}.$$

Модель B:

$$\hat{y}_i = \hat{w}_2^B x_i^{(2)}.$$

Модель C:

$$\hat{y}_i = \hat{w}_1^C x_i^{(1)}.$$

Модель D:

$$\hat{y}_i = 0.$$

Во всех моделях, кроме модели D, веса оцениваются с помощью метода наименьших квадратов (в модели D оценивать нечего). Для каждой из моделей найдите среднее смещение и разброс предсказания по всем четырём объектам. Укажите, какая модель будет давать наименьшую ожидаемую квадратичную ошибку в зависимости от значения σ^2.

Задача 5. (10 баллов) Рассмотрим задачу регрессии. Данные выглядят следующим образом:

$x^{(1)}$	$x^{(2)}$	y
4	0	1
0	-3	5

Используем метод наименьших квадратов с L_1-регуляризатором:

$$\|X\hat{w} - y\|^2 + \|\hat{w}\|_1 \to \min_{\hat{w}}.$$

Будем искать решение этой оптимизационной задачи с помощью градиентного спуска с параметром η (learning rate) равным 1. Найти результат первого шага градиентного спуска из точки $w^{(0)} = (-1, 1)$.

Задача 6. (18 баллов) Влияет ли масштабирование признаков на результат предсказания

a. метода k ближайших соседей;

b. обычной линейной регрессии с МНК-оценкой;

c. линейной регрессии с L_2-регуляризатором?

Если нет, докажите. Если да, приведите пример данных (исходных и перемасштабированных) и точки, для которой получаются разные предсказания.

Формально: пусть исходные признаки записаны виде векторов $x^{(1)}, \ldots, x^{(d)} \in \mathbb{R}^n$ и зафиксирован вектор коэффициентов $c = (c_1, \ldots, c_d)$, все $c_i \neq 0$, $i = 1, \ldots, d$. Новые признаки определяются как $\tilde{x}^{(i)} = c_i x^{(i)}$. Исходные признаки совокупно обозначим через X, новые — через $\tilde{X}$. Вектор ответов обозначим через y. Обозначим через $a(X, y)(x_{new})$ результат предсказания алгоритма, обученного по данным (X, y) на новом объекте $x_{new} = (x_{new}^{(1)}, \ldots, x_{new}^{(d)})$. Верно ли, что для данного алгоритма машинного обучения a для любых данных (X, y), любого набора коэффициентов c и любого нового объекта x_{new},

$$a(X, y)(x_{new}) = a(\tilde{X}, y)(\tilde{x}_{new}),$$

где $\tilde{x}_{new} = (c_1 x_{new}^{(1)}, \ldots, c_d x_{new}^{(d)})$.