

Алгоритмы машинного обучения 1

Задание 1 Дополните предложения

Используйте слова: признаки, классификация, кластер, регрессия.

1. Характеристики объекта, используемые для анализа, — это _____.
2. Группа похожих объектов — это _____.
3. Определение категории объекта — это _____.
4. Прогнозирование числового значения — это _____.

Задание 2 Установите соответствие

Запишите букву подходящего описания в столбце «Ответ».

Алгоритм	Ответ	Описание
1. k-means		А. Прогнозирует число с помощью линейной зависимости.
2. k-NN		Б. Разделяет объекты на группы по сходству признаков.
3. Линейная регрессия		В. Определяет категорию по ближайшим известным примерам.

Задание 3 Верно или неверно?

Поставьте галочку в нужном столбце.

Утверждение	Верно	Неверно
1. В k-means количество кластеров задают заранее.		
2. В k-NN буква k означает количество кластеров.		
3. Линейная регрессия прогнозирует числовое значение.		
4. Для k-means нужны готовые правильные категории объектов.		
5. Ошибки в исходных данных могут повлиять на результат.		

Задание 4 Восстановите порядок действий

Пронумеруйте действия k-means от 1 до 5.

Действие	Номер
Пересчитать центры: найти средние значения признаков объектов каждой группы.	
Задать количество кластеров k.	
Повторять распределение и пересчёт центров до стабилизации результата.	
Выбрать начальные центры кластеров.	
Отнести каждый объект к ближайшему центру.	

Задание 5 Выберите алгоритм

Запишите: k-means, k-NN или линейная регрессия.

Ситуация	Алгоритм
1. Разделить покупателей на группы по числу покупок и сумме расходов. Готовых групп нет.	
2. Определить вид нового фрукта по массе и размеру. Названия фруктов в примерах известны.	
3. Прогнозировать стоимость квартиры по её площади, используя известные площади и цены.	

Объясните выбор для одной ситуации.

Задание 6 Определите результат k-NN

При $k = 5$ выбраны ближайшие соседи нового фрукта:

Сосед	Категория
1	Яблоко
2	Груша
3	Яблоко
4	Яблоко
5	Груша

1. К какой категории отнесут новый фрукт?

2. Почему?

3. Что означает число 5?

Задание 7 Исправьте ошибки

Найдите и исправьте три ошибки в объяснении ученика.

«k-means относится к обучению с учителем: правильная группа каждого объекта известна заранее. В k-NN буква k показывает количество кластеров».

- _____
- _____
- _____

Задание 8 Объясните

А. Масса фруктов указана в граммах, размер — в сантиметрах. Почему для k-NN может потребоваться привести признаки к сопоставимому масштабу?

Б. Модель стоимости доставки создана по данным для расстояний от 1 до 20 км. Надёжен ли её прогноз для 500 км? Почему?