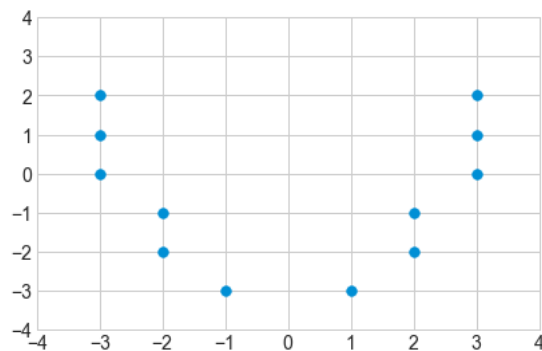


## Самостоятельная работа номер 1. Вариант I

Решите все задания. Все ответы должны быть обоснованы. Все решения должны быть чётко приведены для каждого пункта. Рисунки должны быть чёткими, понятными. Все линии должны быть подписаны. Списывание карается обнулением баллов. Удачи!

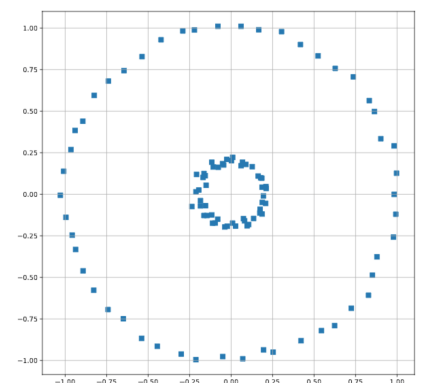
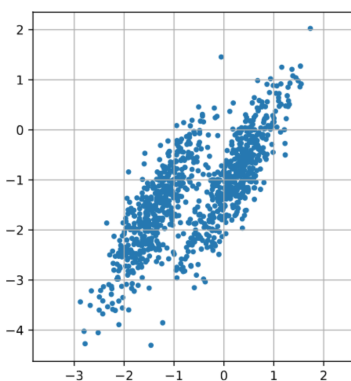
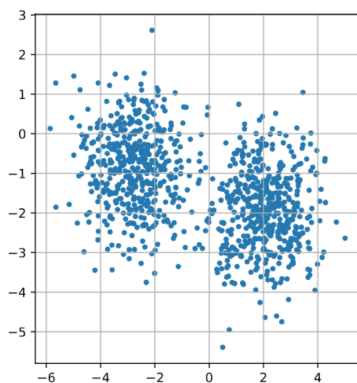
### [7] Задание 1

Примените метод K-means с  $K = 4$ . В качестве начальных точек используйте  $(1, 1)$ ,  $(-1, 1)$ ,  $(0, -2)$ ,  $(0, -3)$ . Опишите алгоритм выбора оптимального числа кластеров для k — means (не надо это проделывать, просто опишите алгоритм).



### [3] Задание 2

Обозначьте расположение центроидов и границ кластеров после применения метода K-means с  $K = 2$  на следующих данных:



## Самостоятельная работа номер 2. Вариант I

Решите все задания. Все ответы должны быть обоснованы. Все решения должны быть чётко приведены для каждого пункта. Рисунки должны быть чёткими, понятными. Все линии должны быть подписаны. Списывание карается обнулением баллов. Удачи!

### [2.5] Задача 1

1. Мы стремимся переобучить любой алгоритм, так как в таком случае он хорошо запомнит выборку.

Да

Нет

2. Коэффициент силуэта используется, чтобы оценить успешность кластеризации и рассчитывается как доля правильно предсказанных кластеров.

Да

Нет

3. Качество модели может сильно зависеть от того как именно мы поделили. выборку на трэйн и тест. Чтобы сгладить этот эффект используют кросс-валидацию.

Да

Нет

4. Метод ближайших соседей никогда не переобучается.

Да

Нет

5. Для задачи кластеризации, в отличие от задачи классификации, мы заранее знаем правильные ответы.

Да

Нет

6. Хорошо, то не было вопросов про случайный лес.

Да

Нет

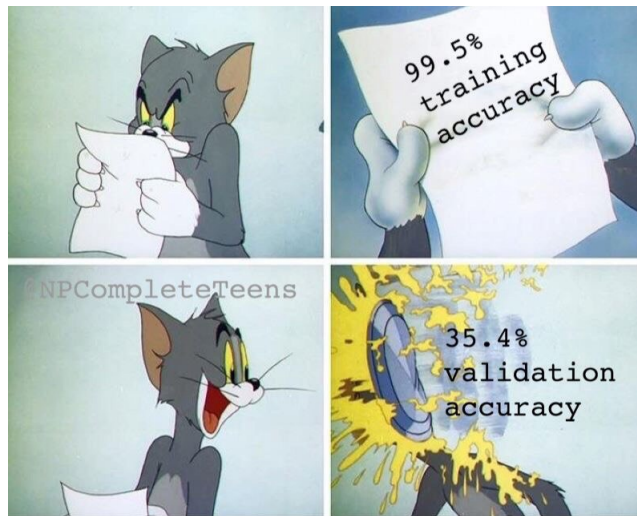
### [6] Задача 2

На плоскости расположены колонии рыжих и чёрных муравьёв. Рыжих колоний три и они имеют координаты  $(-1, 1)$ ,  $(1, -1)$  и  $(1, 1)$ . Чёрных колоний одна и она имеет координаты  $(0, 0)$ .

1. Поделите плоскость на «зоны влияния» рыжих и чёрных используя метод одного и трёх ближайших соседей.
2. С помощью кросс-валидации с выкидыванием отдельных наблюдений выберите оптимальное число соседей  $k$  перебрав  $k \in \{1, 3\}$ . Целевой функцией является количество несовпадающих прогнозов.

### [1.5] Задача 3

Объясните мемас:



## Самостоятельная работа номер 3. Вариант I

Решите все задания. Все ответы должны быть обоснованы. Все решения должны быть чётко приведены для каждого пункта. Рисунки должны быть чёткими, понятными. Все линии должны быть подписаны. Списывание карается обнулением баллов. Удачи!

### [3] Задача 1

Драгомир пытается предсказать продажи видео-игр. Для моделирования он использует две переменные:  $x_1$  - возраст игры,  $x_2$  - на кого она ориентирована. Если на мужчин,  $x_2 = 1$ , если на женщин,  $x_2 = 0$ . Целевая переменная  $y$  - сумма продаж. Драгомир оценил линейную регрессию:

$$y = 1000 - 100 \cdot x_1 + 200 \cdot x_2.$$

Проинтерпретируйте полученные коэффициенты. Предположим, что мы выпускаем на рынок свежую игру для женщин. Спрогнозируйте наши продажи.

### [5] Задача 2

Маше 13 лет. Всю свою жизнь она занималась коллекционированием моделей. Вчера она общалась с Мишей. Он тоже коллекционер. Он спросил у неё, какое у её моделей качество? Маша не смогла ответить и решила проверить его. У неё есть три наблюдения  $y_i$ . Она для каждого построила прогнозы. Найдите для её прогнозов MAE и RMSE.

$y_i$	1	2	3
Нейросеть	2	3	1
Регрессия	2	3	4
Лес	1	1	1

### [2] Задача 3

Объясните мемас:

