

ФИО: _____

Отметьте верные утверждения

Мы хотим посчитать точность **precision**. Что из нижеперечисленного мы должны посчитать чтобы получить результат (минимально, то есть только то что будем использовать):

- FP
- TP
- FN
- TN

Мы хотим посчитать полноту **recall**. Что из нижеперечисленного мы должны посчитать чтобы получить результат (минимально, то есть только то что будем использовать):

- FP
- TP
- FN
- TN

Мы хотим посчитать F-меру **Fscore**. Что из нижеперечисленного мы должны посчитать чтобы получить результат (минимально, то есть только то что будем использовать):

- FP
- TP
- FN
- TN

Если мы поменяем местами позитивный и негативные классы (класс 1 станет классом 0, а класс 0 станет классом 1), что произойдет? Отметьте верные утверждения.

- ассигасу может меняться
- precision может меняться
- recall может меняться

В общем случае нам **нет необходимости** нормализовать фичи для алгоритма

- k-nn алгоритм
- решающее дерево
- случайный лес

Нет возможности построить ROC кривую для алгоритма 1 ближайший сосед для бинарной классификации, потому что алгоритм дает только 2 значения 0 и 1 и нет возможности использовать трешхолд.

Отметьте верные утверждения для алгоритма k ближайших соседей.

- KNN используется только для классификации.
- Обучение KNN это просто запоминание объектов (характеристик и меток ответов)
- В общем случае необходимо нормализовать данные.
- Использовать 4 ближайших соседа для бинарной классификации в алгоритме knn не очень хорошо (будут проблемы). Напишите почему?

Отметьте алгоритм который **может дать** неправильный ответ на объекте из тренировочной выборки (на объекте на котором мы тренировали алгоритм)

- knn для любого $k > 1$
- 1 ближайший сосед
- решающее дерево с максимальной глубиной = 5
- случайный лес

Для отрисовки *ROC* мы используем true positive rate(*TPR*) и false positive rate(*FPR*). Напишите формулы для этих величин используя только *TP*, *FP*, *FN*, *TN*:

• $TPR =$ _____

• $FPR =$ _____

Рассмотрим алгоритм который решает задачу бинарной классификации и дает ответы на входной текст "spam" или "not spam". Если мы увеличим порог, что **может** произойти:

- Recall может не измениться.
- Recall может возрасти.
- Recall может уменьшиться.
- Precision может не измениться.
- Precision может увеличиться.
- Precision может уменьшиться.

Представим что у нас есть идеальный классификатор (качество 100%) натренированный на логической регрессии. Значит ли это что мы обязательно получим идеальный классификатор если натренируем на тех же данных:

- knn.
- метод опорных векторов с параметром $C = 0$, и без ядрового преобразования.
- решающее дерево.

Мы используем иерархическую кластеризацию. Построили дендрограмму. Может ли случиться ситуация, что мы не можем **однозначно** раскластеризовать нашу выборку на n классов, учитывая что всего объектов было больше чем n

- Да
 - Нет
- почему?

отметьте верные утверждения об ансамблях

- В случайном лесе мы можем тренировать все деревья параллельно
- В градиентном бустинге мы можем тренировать все алгоритмы параллельно
- В стекинге мы можем тренировать все алгоритмы параллельно

Отметьте те алгоритмы для которых нам необходимо зафиксировать метрику:

- random forest
- наивный байес
- knn
- k means
- решающее дерево

Если мы используем линейную регрессию на выборке с 10 осмысленными признаками, то сколько параметров нам надо будет обучить?

- 10
- 9
- 11
- 22
- 20