



## ТЕСТ МОЖНО СДАТЬ ТОЛЬКО 1 РАЗ, НАЖАВ НА КНОПКУ "Сохранить решение"

В вопросе может быть **несколько** вариантов правильного ответа (то есть от 1 до кол-ва ответов в вопросе). Вопрос засчитывается, если выбраны **ВСЕ** правильные варианты и **НЕ ВЫБРАНЫ** ВСЕ неправильные варианты.

В тестовых заданиях первая галочка — правильный ответ, вторая галочка — выбранный ответ. Цвет обозначает, правильно ли в данном пункте поставлена галочка. Если все пункты верные (галочки совпадают / все пункты зеленые), то за задание ставится полный балл, в противном случае ставится 0 баллов.

1. Пусть  $H(Y)$  - энтропия сл. величины  $Y$ ,  $H(X,Y)$  - энтропия пары сл. величин. Выберите верные определения взаимной информации (mutual information)

☒ ☒  $H(X)+H(Y)-H(X,Y)$

☐ ☒  $H(X,Y)-H(X)-H(Y)$

**Балл: 0**

**Комментарий к правильному ответу:**

2. Выберите методы, которые сами отбирают признаки, без подключения отдельного модуля отбора признаков:

☐ ☐ метод К ближайших соседей

☐ ☐ линейная регрессия с L2 регуляризацией

☐ ☐ регрессия Надарая-Ватсона

☐ ☐ линейный классификатор без регуляризации

☒ ☒ линейная регрессия с L1 регуляризацией

☐ ☐ линейная регрессия без регуляризации

☒ ☒ решающее дерево

☐ ☐ линейный классификатор с L2 регуляризацией

☒ ☐ линейный классификатор с L1 регуляризацией

☒ ☒ случайный лес

**Балл: 0**

**Комментарий к правильному ответу:**



3. Возможна ли ситуация, когда признак и отклик-зависимые случайные величины, но при добавлении этого признака в число других признаков и повторного обучения модели точность модели не изменяется?

☐ ☒ нет

☒ ☐ да

**Балл: 0**

**Комментарий к правильному ответу:**

Например, когда этот признак продублирован дважды и присутствует среди других признаков

4. Может ли добавление нового признака и последующее обучение модели с ним снизить точность на *тестовой* выборке по сравнению с точностью модели без этого признака?

☒ ☐ да

☐ ☒ нет

**Балл: 0**

**Комментарий к правильному ответу:**

5. В генетическом алгоритме какие наборы признаков следует переносить в следующий раунд модификаций без изменения?

☐ ☒ показавшие наихудшее качество

☒ ☐ показавшие наилучшее качество

**Балл: 0**

**Комментарий к правильному ответу:**

6. Рассмотрим оценку значимости признаков с помощью корреляции признака и отклика. Отбор признаков осуществляем выбором максимально скоррелированных признаков. Недостатком такого подхода, по сравнению с перебором всех подмножеств признаков и оценки качества прогноза на них, будет

☒ ☒ отсутствие учета взаимного влияния признаков на отклик, вклад признаков учитывается только по отдельности

☐ ☐ вычислительная неэффективность



- ☒ ☐ учет только линейных зависимостей между признаками и откликом, а не потенциальных нелинейных
- ☒ ☐ потенциальное включение большого числа скоррелированных между собой признаков

**Балл: 0**

**Комментарий к правильному ответу:**

7. Может ли добавление нового признака и последующее обучение модели с ним снизить точность на *обучающей* выборке по сравнению с точностью модели без этого признака?

☒ ☒ да

☐ ☐ нет

**Балл: 2.0**

**Комментарий к правильному ответу:**

8. В генетическом алгоритме признаки, которые часто встречаются в отобранных лучших наборах признаков следует

☐ ☐ мутировать чаще

☒ ☒ мутировать реже

**Балл: 2.0**

**Комментарий к правильному ответу:**