

**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. Рассмотрим метод последовательного жадного включения признаков по одному среди D признаков. Допустим этот метод остановился после выдачи подмножества из K элементов. Сколько раз для этого методу пришлось вычислить критерий качества (или функцию потерь)?

☐ ☐ $K(D-K/2)$

☐ ☐ $K(D-(K+1)/2)$

☒ ☒ $K(D-(K-1)/2)$

Балл: 2.0

Комментарий к правильному ответу:

$$D+(D-1)+\dots+(D-K+1)=DK-(1+2+\dots+(K-1))=DK-(K-1)K/2=K(D-(K-1)/2)$$

2. Ранговая корреляция Спирмена по сравнению с обычной корреляцией между признаком и откликом позволяет

☐ ☐ детектировать любые виды зависимости

☒ ☒ детектировать монотонные виды зависимости

☐ ☐ детектировать только линейные виды зависимости

Балл: 2.0

Комментарий к правильному ответу:

3. 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



- ☐ ☐ вычислительная неэффективность
- ☒ ☒ потенциальное включение большого числа скоррелированных между собой признаков
- ☒ ☒ учет только линейных зависимостей между признаками и откликом, а не потенциальных нелинейных
- ☒ ☐ отсутствие учета взаимного влияния признаков на отклик, вклад признаков учитывается только по отдельности

Балл: 0

Комментарий к правильному ответу:

4. Рассмотрим непрерывную случайную величину X с четной функцией распределения $p(-x)=p(x)$ и другую случайную величину $Z=X^2$. Выберите верное утверждение:

- ☒ ☒ X и Z - зависимые сл. величины, корреляция между ними равна 0.
- ☐ ☐ X и Z - зависимые сл. величины, корреляция между ними отлична от нуля, но занижает реальную степень взаимосвязи между ними.
- ☐ ☐ X и Z - зависимые сл. величины, корреляция между ними равна 1.
- ☐ ☐ X и Z - независимые сл. величины

Балл: 2.0

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: 2.0

Комментарий к правильному ответу:

6. Пусть $H(Y)$ - энтропия сл. величины Y , $H(Y|X)$ - условная энтропия Y при условии др. случайной величины X . Выберите верные определения взаимной информации (mutual information)

- ☒ ☒ $H(Y)-H(Y|X)$



☒ ☒ $H(X) - H(X|Y)$

Балл: 2.0

Комментарий к правильному ответу:

7. Допустим, мы знаем истинные распределения признака отклика. Выберите меры взаимосвязи признака и отклика, которая извлечет любые виды взаимосвязи между признаком и откликом (взаимосвязь на событиях с вероятностью 0 не учитываем):

☐ ☐ ранговая корреляция Спирмена

☐ ☐ корреляция

☒ ☒ взаимная информация

Балл: 2.0

Комментарий к правильному ответу:

8. 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?

☒ ☒ да

☐ ☐ нет

Балл: 2.0

Комментарий к правильному ответу:

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