

Вопрос 1

Выполнен

Баллов: 1,00 из 1,00

Развёрнутый вопрос

Такие задачи иногда будут появляться, чтобы помочь Вам глубже понять материал.

Качество классификации может быть измерено с помощью нескольких метрик. Одни из самых популярных – точность, полнота и F1-мера (ф-мера). В основе этих метрик лежат частоты ошибок первого и второго рода (ложноотрицательных и ложноположительных срабатываний)

	Истинная метка - 0	Истинная метка - 1
Предсказание - 0	Истинно отрицательное предсказание True Negative, TN	Ложно-отрицательное предсказание False Negative, FN
Предсказание - 1	Ложно-положительное предсказание False Positive, FP	Истинно положительное предсказание True Positive, TP

Сокращения TN, FN, FP, TP можно использовать для обозначения количества истинно-отрицательных, ложно-отрицательных, ложноположительных и истинно-положительных предсказаний оцениваемого классификатора на валидационной выборке.

Используя эти обозначения, можно определить точность, полноту и F1-меру следующим образом:

$$Precision = \frac{TP}{TP+FP+\epsilon}, Recall = \frac{TP}{TP+FN+\epsilon}, F_1 = \frac{2PrecisionRecall}{Precision+Recall+\epsilon},$$

где ϵ - малое по модулю положительное число (чтобы не было деления на 0).

Кроме того, иногда используют долю верных угадываний: $Accuracy = \frac{TP+TN}{TN+FN+FP+TP}$

Проанализируйте эти формулы и отметьте в списке ниже истинные утверждения (одно или несколько), касающиеся этих метрик.

При оценке истинности утверждений предполагайте, что количество положительных и отрицательных примеров в валидационной выборке одинаковое, если в варианте ответа не указано другое.

Выберите один или несколько вариантов ответа.

- ☒ a. Если классификатор всегда возвращает 1, полнота на валидационной выборке приблизительно равна 1
- ☐ b. Если классификатор всегда возвращает 0, F1 на валидационной выборке приблизительно равна 1
- ☒ c. Если классификатор всегда возвращает 1, Ассигасу на валидационной выборке равна 0.5 (при этом классы сбалансированы)
- ☐ d. Если классификатор всегда возвращает 1, точность на валидационной выборке равна 0
- ☐ e. Если классификатор всегда возвращает 1, F1 на валидационной выборке приблизительно равна 1
- ☒ f. Если классификатор всегда возвращает 1, F1 на валидационной выборке равна 0.667
- ☐ g. F1 достигает наибольшего значения, когда полнота больше точности
- ☒ h. Если классификатор всегда возвращает 0, Ассигасу на валидационной выборке равна 0.5 (при этом классы сбалансированы)
- ☒ i. Если классификатор всегда возвращает 0, полнота на валидационной выборке равна 0
- ☐ j. F1 достигает наибольшего значения, когда точность больше полноты
- ☒ k. Если классификатор всегда возвращает 0, F1 на валидационной выборке равна 0
- ☐ l. Ассигасу всегда показывает надёжную оценку качества классификатора, даже на выборке со скошенным распределением классов (например, когда количество отрицательных примеров сильно больше количества положительных)
- ☐ m. Если классификатор всегда возвращает 0, точность на валидационной выборке приблизительно равна 1

Ваш ответ верный.

Вопрос 2

Выполнен

Баллов: 1,00 из 1,00

Развёрнутый вопрос

Процесс решения задачи поиска часто состоит из двух крупных шагов, каждый из которых разбивается на более мелкие:

1. Настройка поиска
 1. преобразование объектов (например, текстов) в вещественные вектора
 2. построение поискового индекса
 3. настройка функции ранжирования
2. Выполнение поиска
 1. преобразование запроса в вещественный вектор
 2. грубая выборка кандидатов
 3. сортировка кандидатов с помощью функции ранжирования

Поисковый индекс - набор специальных структур данных, ускоряющих процесс поиска. Так как процесс индексации тоже требует времени, поисковые индексы не всегда имеет смысл строить - например, когда данных мало или данные часто меняются и индекс устаревает быстрее, чем может быть перестроен.

Настройка функции ранжирования выполняется с помощью набора примеров вида "запрос - документ - оценка релевантности по мнению человека". Релевантность - численная величина, характеризующая соответствие найденного документа запросу (чем больше, тем лучше документ подходит под запрос).

Пусть у нас есть коллекция документов, которые на шаге 1.1 были преобразованы в следующие вектора

ID документа	Признаки
1	(0, 1)
2	(1, 0)
3	(1, 0.5)

Также у нас есть функция ранжирования $Relevance(q, d) = -(q_1 - d_1)^2 - 2(q_2 - d_2)^2$, где (q_1, q_2) - признаки запроса, а (d_1, d_2) - признаки документа.

Нам пришел запрос с признаками (1, 1).

Отсортируйте документы в порядке убывания релевантности относительно данного запроса.

3

1

2

Ваш ответ верный.

Вопрос 3

Выполнен

Баллов: 1,00 из 1,00

Выберите истинные утверждения (одно или несколько) касательно задачи поиска текстов.

- ☒ a. Градиентный бустинг может помочь лучше искать документы, когда у нас есть информация о поведении пользователя (задача learning to rank).
- ☐ b. Линейные модели чаще обеспечивают большую полноту по сравнению с дистрибутивно-семантическими моделями.
- ☐ c. Поиск текстов - это задача определения тематики документа.
- ☐ d. В поиске похожих документов в качестве запроса выступает картинка.
- ☐ e. Информация о поведении пользователей доступна с самого начала работы поисковой системы.
- ☒ f. Линейные модели чаще обеспечивают большую точность по сравнению с тематическим моделированием.
- ☒ g. Поиск текстов - это задача получения выборки текстов, релевантных запросу.

Ваш ответ верный.

Вопрос Инфо

Вопрос **4**

Выполнен

Баллов: 1,00 из 1,00

Выберите верные утверждения (одно или несколько) касательно извлечения структурированной информации.

Выберите один или несколько ответов:

- ☐ а. Для извлечения информации можно использовать тематическое моделирование
- ☒ b. Извлечение именованных сущностей сводится к задаче классификации токенов
- ☒ с. Для извлечения информации можно использовать нейросети

Ваш ответ верный.

Вопрос **Инфо**