

# MADE: Machine Learning. HA №1.

Deadline: 19/10/2020 23:59 Moscow

Time.

Всего 2/13 ?

Важно указать тот email адрес, который вы использовали для регистрации на портале MADE, а также настоящие ФИО. По этим данным результаты работы будут полуавтоматически перенесены на портал.

Адрес электронной почты \*

iewfor@aol.com

Баллов: 0 из 0.

Фамилия \*

Wmf7VnYHxhlvr

Имя \*

QLScvZ935nv

Отчество

rHBcqXdYieN

Базовые концепты машинного обучения

Баллов: 1 из 6.

✗ Выберите из списка задачи, относящиеся к классификации

0 из 1

- ☒ Прогнозирование оттока: для каждого клиента требуется определить, с какой вероятностью он покинет компанию через выбранный промежуток времени в будущем ✓
- ☒ Прогноз спроса на товары и услуги: для каждого товара/услуги требуется спрогнозировать спрос с фиксированным горизонтом в будущем ✗
- ☒ Сегментация пользователей: требуется проанализировать аудиторию и разделить её на сегменты (описание сегментов и их количество заранее неизвестны) ✗
- ☒ Товарные рекомендации: для каждого пользователя сайта магазина техники требуется построить персональные рекомендации товаров магазина ✗
- ☒ Прогноз спроса на наличные в АТМ банка: для каждого АТМ требуется спрогнозировать сколько наличных будет запрошено пользователями на фиксированном горизонте в будущем ✗
- ☒ Автоматическое определение жанра фильма по текстовой аннотации ✓
- ☒ Анализ эмоциональной окраски сообщений в службе поддержки: определить носит ли сообщение позитивный, негативный или нейтральный характер ✓

Правильный ответ

- ☒ Прогнозирование оттока: для каждого клиента требуется определить, с какой вероятностью он покинет компанию через выбранный промежуток времени в будущем
- ☒ Автоматическое определение жанра фильма по текстовой аннотации
- ☒ Анализ эмоциональной окраски сообщений в службе поддержки: определить носит ли сообщение позитивный, негативный или нейтральный характер

✗ Выберите задачи из списка, относящиеся к регрессии

0 из 1

- ☒ Прогнозирование оттока: для каждого клиента требуется определить, с какой вероятностью он покинет компанию через выбранный промежуток времени в будущем ✗
- ☒ Прогноз спроса на товары и услуги: для каждого товара/услуги требуется спрогнозировать спрос с фиксированным горизонтом в будущем ✓
- ☒ Сегментация пользователей: требуется проанализировать аудиторию и разделить её на сегменты (описание сегментов и их количество заранее неизвестны) ✗
- ☒ Детектирование производственного брака: требуется на ранней стадии производства выявить продукты, которые в результате могут оказаться бракованными ✗
- ☒ Прогноз спроса на наличные в АТМ банка: для каждого АТМ требуется спрогнозировать сколько наличных будет запрошено пользователями на фиксированном горизонте в будущем ✓
- ☒ Определить жанр фильма по текстовой аннотации ✗
- ☒ Анализ эмоциональной окраски сообщения в службе поддержки: определить носит ли сообщение позитивный, негативный или нейтральный характер ✗

Правильный ответ

- ☒ Прогноз спроса на товары и услуги: для каждого товара/услуги требуется спрогнозировать спрос с фиксированным горизонтом в будущем
- ☒ Прогноз спроса на наличные в АТМ банка: для каждого АТМ требуется спрогнозировать сколько наличных будет запрошено пользователями на фиксированном горизонте в будущем

✗ Задача ранжирования отличается от задачи регрессии следующим (укажите все правильные утверждения):

0 из 1

- ☒ в задаче ранжирования относительный порядок числовых ответов важнее, чем абсолютное значение числового ответа, потому что ответ будет использоваться для упорядочивания объектов друг относительно друга ✓
- ☒ в задаче ранжирования числовые ответы должны быть исключительно в интервале от 0 до 1 ✗
- ☒ в задаче ранжирования запрещено иметь одинаковые ответы на разных объектах, а в задаче регрессии это разрешено ✗
- ☒ если умножить все ответы модели на 10, то в случае задачи ранжирования качество решения не ухудшится, а в случае задачи регрессии ухудшится существенно ✓
- ☒ если умножить все ответы модели на 10, то в случае задачи регрессии качество решения не ухудшится, а в случае задачи ранжирования ухудшится существенно ✗

Правильный ответ

- ☒ в задаче ранжирования относительный порядок числовых ответов важнее, чем абсолютное значение числового ответа, потому что ответ будет использоваться для упорядочивания объектов друг относительно друга
- ☒ если умножить все ответы модели на 10, то в случае задачи ранжирования качество решения не ухудшится, а в случае задачи регрессии ухудшится существенно

✗ При построении модели и оценке её качества, данные разделяют на обучающую, валидационную и тестовую выборку для того, чтобы (укажите все правильные утверждения)

0 из 1

- ☒ Усложнить процесс построения модели. Достаточно деления данных на обучение и тест, выделение валидационной выборки никаких существенных плюсов не даёт. ✗
- ☒ Ускорить процесс обучения модели за счет того, что для обучения используется только обучающая выборка и она меньше по объему, чем все исходные данные ✗
- ☒ Более точно оценить качество работы модели за счет того, что будет сделана оценка качества не только на тех данных, которые были продемонстрированы модели в процессе обучения, но и на тех данных, которые являются новыми для модели. ✓
- ☒ Данные разделяют по принципу: наиболее качественные наблюдения отбираются в обучающую выборку, нетипичные объекты отправляют в валидационную выборку, а шумные наблюдения и выбросы идут в тест. За счет такой "фильтрации" данных для обучения достигается рост качества модели, а также валидация по нетипичным объектам позволяет получить надежную модель ✗
- ☒ Снизить вероятность слишком тонкой подстройки параметров модели под фиксированную выборку: параметры подбираются по валидации, но модель дополнительно оценивается по тесту. ✓

Правильный ответ

- ☒ Более точно оценить качество работы модели за счет того, что будет сделана оценка качества не только на тех данных, которые были продемонстрированы модели в процессе обучения, но и на тех данных, которые являются новыми для модели.
- ☒ Снизить вероятность слишком тонкой подстройки параметров модели под фиксированную выборку: параметры подбираются по валидации, но модель дополнительно оценивается по тесту.

✗ При подборе гиперпараметров моделей машинного обучения требуется проведение кросс-валидации. Какие из перечисленных ниже стратегий кросс-валидации применимы для оптимизации гиперпараметров модели?

0 из 1

- ☒ стратегия k-fold ✓
- ☒ стратегия leave-one-out (loo) ✓
- ☒ стратегия tk-fold ✓
- ☒ стратегия stratified loo ✗
- ☒ стратегия k random splits ✓
- ☒ стратегия k-loo ✗
- ☒ стратегия stratified tk-fold ✓

Правильный ответ

- ☒ стратегия k-fold
- ☒ стратегия leave-one-out (loo)
- ☒ стратегия tk-fold
- ☒ стратегия k random splits
- ☒ стратегия stratified tk-fold

✓ Каким образом строится финальная модель после подбора параметров методом кросс-валидация?

1 из 1

- ☐ Каждый набор параметров тестируется на своем фолде, оценивается качество. Значения параметров усредняются с весами, пропорциональными качеству на фолдах. Именно с этими значениями параметров обучается финальная модель.
- ☐ Каждый набор параметров тестируется на своем фолде. Для финальной модели выбираются параметры с фолда, показавшего лучший результат по метрике. Далее с этими параметрами строится финальная модель.
- ☒ Каждый набор параметров тестируется на k фолдах, после чего оценка качества для каждого набора параметров усредняется по фолдам. Для финальной модели выбираются параметры, показавшие лучшее среднее на фолдах. Далее с этими параметрами строится финальная модель. ✓
- ☐ Каждый набор параметров тестируется на k фолдах, после чего оценка качества для каждого набора параметров усредняется по фолдам. Значения параметров усредняются с весами, пропорциональными качеству на фолдах. Именно с этими значениями параметров обучается финальная модель.
- ☐ Каждый набор параметров тестируется на своем фолде, оценивается качество. Значения параметров усредняются. Именно с этими значениями параметров обучается финальная модель.

Простые алгоритмы обучения с учителем

Баллов: 0 из 3.

✗ Вы решаете задачу бинарной классификации с помощью дерева решений. Ваш набор данных включает 10000 наблюдений с 10 бинарными признаками. Каково минимальное возможное число листьев в дереве решений?

0 из 1

- ✗

Правильные ответы

- 1
- 2
- 0

Комментарий

В случае, если в выборке присутствуют экземпляры только одного класса, то лист будет всего один. Однако, ответ 2 также является верным, потому что из-за формулировки "задача бинарной классификации" можно предположить, что в выборке есть хотя бы по 1 представителю каждого класса.

✗ Вы решаете задачу бинарной классификации с помощью дерева решений. Ваш набор данных включает 10000 наблюдений с 10 бинарными признаками. Каково максимально возможное число листьев в дереве решений?

0 из 1

- ✗

Правильные ответы

- 1024
- 2\*10

✗ В чем состоит предположение Наивного Байесовского классификатора?

0 из 1

- ☒ Признаки объектов распределены нормально ✗
- ☐ Признаки объектов распределены нормально и независимы
- ☐ Признаки объектов независимы
- ☐ Признаки объектов распределены стандартно нормально

Правильный ответ

- ☒ Признаки объектов независимы

Валидация моделей

Баллов: 1 из 4.

✗ Вы решаете задачу бинарной классификации. В качестве baseline вы выбрали алгоритм, который вместо прогноза уверенности в том, что объект принадлежит классу 1, выдает случайные действительные ответы, равномерно распределенные от 0 до 1. Каким будет ROC AUC для этих прогнозов?

0 из 1

- ✗

Правильные ответы

- 0.5
- 0.5
- 0.5
- 0.5

✗ Вы решаете задачу бинарной классификации. В качестве baseline вы выбрали алгоритм, который вместо прогноза уверенности в том, что объект принадлежит классу 1, выдает на любом объекте ответ 1/n, где n - размер выборки. Каким будет ROC AUC для этих прогнозов?

0 из 1

- ✗

Правильные ответы

- 0.5
- 0.5
- 0.5
- 0.5

✓ Вы решаете задачу ранжирования. Вы обучили модель ранжирования и оценили её качество с помощью метрики ROC AUC, получили значение A1. Как изменится ROC AUC (назовём его значение A2) модели, если вы умножите значение каждого из прогнозов на 2?

1 из 1

- ☐ A2 = 0.5\*A1
- ☐ A2 = 2\*A1
- ☐ A2 = 0
- ☒ A2 = A1 ✓
- ☐ A2 > A1
- ☐ A2 < A1

✗ Выберите все верные утверждения про метрики качества аккуратность (accuracy), точность (precision) и полнота (recall) в задаче классификации

0 из 1

- ☒ Значение метрик точность (precision) и полнота (recall) всегда совпадают ✗
- ☒ Значение метрик точность (precision) и аккуратность (accuracy) совпадают в случае бинарной классификации с соотношениями классов 50/50 ✗
- ☒ Метрика точность (precision) оценивает долю объектов, верно отнесенных к целевому классу, среди всех объектов, которые модель отнесла к целевому классу ✓
- ☒ Метрика полнота (recall) оценивает долю объектов, верно отнесенных к целевому классу, среди всех объектов, которые модель отнесла к целевому классу ✗
- ☒ Метрика точность (precision) оценивает долю объектов, верно отнесенных к целевому классу, среди всех объектов целевого класса в выборке ✗
- ☒ Метрика полнота (recall) оценивает долю объектов, верно отнесенных к целевому классу, среди всех объектов целевого класса в выборке ✓
- ☒ Метрики точность (precision) и полнота (recall) в сумме дают единицу ✗

Правильный ответ

- ☒ Метрика точность (precision) оценивает долю объектов, верно отнесенных к целевому классу, среди всех объектов, которые модель отнесла к целевому классу
- ☒ Метрика полнота (recall) оценивает долю объектов, верно отнесенных к целевому классу, среди всех объектов целевого класса в выборке

Компания Google не имеет никакого отношения к этому контенту. - Условия использования - Политика конфиденциальности

Google Формы