УТВЕРЖДАЮ

Зав. кафедрой СП ЮУрГУ

\_\_\_\_\_\_\_\_\_Л.Б. Соколинский

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г.

Фонд оценочных средств ООП «Искусственный интеллект и инженерия данных» по направлению 09.04.04 – Программная инженерия

Дисциплина «Основы машинного обучения»

| **№ КМ** | **Вид КМ** | **Наименование КМ** | **Оценочные средства** |
| --- | --- | --- | --- |
|  | Текущий контроль | Тест 1 | Пример теста:  1)    2)    3) |
|  | Текущий контроль | Тест 2 | Пример теста:  1)    2)    3) |
|  | Текущий контроль | Тест 3 | Пример теста:  1)    2)    3) |
|  | Текущий контроль | Тест 4 | Пример теста:  1)    2)    3) |
|  | Текущий контроль | Тест 5 | 1)    2)    3) |
|  | Текущий контроль | Тест 6 | 1)    2)    3) |
|  | Текущий контроль | Тест 7 | 1)    2)    3) |
|  | Текущий контроль | Практика 1 | Вопросы для подготовки к устному опросу:  1. Алгоритм одномерной регрессии  2. Алгоритм множественной регрессии |
|  | Текущий контроль | Практика 2 | Вопросы для подготовки к устному опросу:  1. Функция сигмоиды  2. Метод максимального пдавдоподобия  3. Явление недообучения  4. Явление переобучения  5. Регуляризация |
|  | Текущий контроль | Практика 3 | Вопросы для подготовки к устному опросу:  1. Трудоразделимые классы  2. Зависимость точности классификатора от k в методе kNN  3. Дерево решений  4. Ансамбль деревьев – случайный лес  5. Градиентный бустинг |
|  | Текущий контроль | Практика 4 | Вопросы для подготовки к устному опросу:  1. Линейный SVM  2. Нелинейный SVM  3. Зависимость классификатора от его параметров |
|  | Текущий контроль | Практика 5 | Вопросы для подготовки к устному опросу:  1. Понижение размерности данных до 2  2. Разделимость классов для понижения размерности  3. Метод главных компонент  4. Метод t-SNE  5. Валидация модели машинного обучения  6. Визуализация данных для работы |
|  | Текущий контроль | Практика 6 | Вопросы для подготовки к устному опросу:  1. Метрики оценки качества моделей  2. Отбор моделей  3. Корректировка параметров |
|  | Промежуточный контроль | Итоговый тест | Вопросы для подготовки к экзамену по дисциплине:  1. Сильный и слабый искусственный интеллект  2. Линейная регрессия  3. Логистическая регрессия  4. Ансамбли деревьев решения  5. Метод kNN  6. Метод CV, GridSearch  7. Метод SVM  8. Использование PSA для уменьшения размерности данных  9. Метод t-SNE  10. Валидация модели машинного обучения  11. Визуализация данных |

Паспорт фонда оценочных средств приведен в п. 6.3 РПД.

Разработчик О.Н. Иванова

ФГАОУ ВО «Южно-Уральский государственный университет   
(национальный исследовательский университет)»

Кафедра системного программирования

Дисциплина «Основы машинного обучения»

ИТОГОВЫЙ ТЕСТ

| № | Вопрос | Варианты ответа |
| --- | --- | --- |
|  | Метод градиентного спуска... | уже не используется а машинном обучении  как правило, требует нормализации данных  это не итерационный алгоритм  не требует выбора шага обучения |
|  | Какая из перечисленных задач не является задачей линейной регресии? | Определение сорта ириса  Определение цены на квартиру по ее площади  Прогнозирование температуры на завтра  Прогнозирование дохода ресторана в зависимости от количества населения в городе |
|  | Метод решения задачи линейной регрессии, реализованный в sklearn, называется... | метод выпуклой оптимизации  метод градиентного спуска  MinMaxScaler  normal equation |
|  | Для минимизации целевой функции в задаче линейной регрессии используется метод... | главных комопонент  опорных векторов  градиентного спуска  ближайших соседей |
|  | От скольких переменных зависит целевая функция в задаче одномерной линейной регрессии? | n+1, где n - число обучающих элементов  n, где n - число обучающих элементов  1  2 |
|  | При решении задачи линейной регрессии методом градиентного спуска, мы | минимизируем функцию h(x)=θ0+θ1x  увеличиваем вероятность застрять в локальном минимуме  минимизируем целевую функцию |
|  | Сколько неизвестных параметров θθ ищется методом градиентного спуска в задаче множественной линейной регрессии с nn входными признаками? | n  2  n−1  n+1 |
|  | Что из ниже перечисленных моделей более подвержена явлению переобучения? | гребневая регрессия  лассо  ElasticNet  линейная регрессия |
|  | Решается задача классификации ирисов Фишера. Какой из нижеприведенных методов НЕ подходит для ее решения? | деревья решения  градиентный бустинг  метод ближайших соседей  метод опорных векторов  линейная регрессия  логистическая регрессия  взвешенный метод ближайших соседей |
|  | Какой из нижеприведенных приемов НЕ является способом борьбы с переобучением? | регуляризация  нормализация признаков  понижение порядка модели  уменьшение числа признаков |
|  | При построении ансамбля методом градиентного бустинга образцы, которые доставляют большую ошибку первой обученной модели | имеют больший шанс попасть в выборку для обучения второй модели  исключаются из обучающей выборки для последующих моделей  ничем не отличаются от остальных и могут также случайно быть выбраны |
|  | Какой из следующих методов наименее чувствителен к отсутствию предварительной нормализации признаков? | метод логистической регрессии  деревья решений  метод линейной регрессии  метод ближайших соседей |
|  | В какой из этих моделей увеличение числа деревьев дает, как правило, лучший эффект? | градиентный бустинг  случайный лес |
|  | Чему равно оптимальное число ближайших соседей в методе kNN? | равно n+1, где n - число признаков  подбирается экспериментально в каждой задаче  равно 1  равно числу признаков n  равно 10 |
|  | Что не является параметром метода kNN? | функция расстояния  шаг обучения  число соседей  веса соседей |
|  | Что из нижеперечисленного является функцией потерь в методе kNN? | нет функции потерь  логлосс  MSE  RMSE  hinge |
|  | Большое значение параметра С в методе опорных векторов приводит к тому, что | модель имеет более широкий зазор  модель склонна к недообучению  модель склонна к переобучению и ловит выбросы  более гладкой границе решений |
|  | Kernel trick или введение новых признаков F вместо исходных признаков X каждого объекта нужно | для решения задачи регрессии с помощью SVM  для максимизации зазора  для ускорения сходимости SVM  для создания нелинейного классификатора SVM |
|  | При использовании метода SVM | нужно всегда использовать kernel trick и вводить новые признаки  решается задача выпуклой оптимизации и мы всегда находим глобальный минимум  мы можем застрясть в локальном минимуме |
|  | Можно ли использовать метод PCA с числом главных компонент равным числу признаков? | да, при этом уберем корреляцию между признаками  да, но при этом входные данные не будут никак преобразованы  нет, в этом случае алгоритм разойдется  нет, в этом случае матрица сингулярного разложения будет вырождена |
|  | Укажите смысл параметра перплексити метода t-SNE | число главных компонент  шаг обучения  число близких соседей  пороговое значение дисперсии |
|  | Что понимается под проблемой скученности? | a.  невозможность сохранить все расстояния между соседями при отображении в пространство меньшей размерности    b.  большой объем выборки и, как следствие, много похожих элементов    c.  вырожденность сингулярной матрицы при применении PCA к выборкам большой размерности |
|  | Как называется метод тестирования точности модели, при котором все имеющиеся данные делятся на два подмножества, на одном из которых модель обучается, на другом тестируется. | K-fold-cross-validation  held-out/hold-out set  MinMaxScaler  Leave-one-out cross validation |
|  | Решетчатый поиск используется для | нормализации данных  оценки точности модели  перебора возможных значений параметров |
|  | Какой из следующих способов контроля качества обучения будет самым затратным в вычислительном плане? | Leave-one-out cross validation  held out/hold out set  K-fold cross validation |