

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Физико-технический факультет

УТВЕРЖДЕНО:

Декан

Ю.Н. Рыжих

Оценочные материалы по дисциплине

Введение в машинное обучение

по направлению подготовки

15.03.06 Мехатроника и робототехника

Направленность (профиль) подготовки:
Промышленная и специальная робототехника

Форма обучения

Очная

Квалификация

Инженер, инженер-разработчик

Год приема

2024

СОГЛАСОВАНО:

Руководитель ОПОП

Е.И. Борзенко

Председатель УМК

В.А. Скрипняк

Томск – 2024

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен выявлять естественнонаучную сущность проблем, возникающих в ходе профессиональной деятельности, привлекать для их решения физико-математический аппарат и современные компьютерные технологии;

ОПК-4 Способен осуществлять контроль соответствия разрабатываемых проектов и технической документации стандартам, техническим условиям и другим нормативным документам;

ОПК-7 Способен нести ответственность за принятие решений по части или всем сложным видам инженерной деятельности;

ОПК-8 Способен понимать принципы работы современных информационных технологий, обрабатывать и использовать новую информацию в своей предметной области, предлагать новые идеи и подходы к решению инженерных задач;

ПК-1 Способность составлять математические модели мехатронных и робототехнических систем, их подсистем и отдельных элементов и модулей, включая информационные, электромеханические, гидравлические, электрогидравлические, электронные устройства и средства вычислительной техники;

ПК-3 Способность проводить вычислительные эксперименты с использованием стандартных программных пакетов с целью исследования математических моделей мехатронных и робототехнических систем.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

РООПК-2.1 Знает методику выявления естественнонаучной сущности проблем, возникающих в ходе профессиональной деятельности, и методику привлечения физико-математического аппарата и современные компьютерных технологий для их решения

РООПК-2.2 Умеет выявлять естественнонаучную сущность проблем, возникающих в ходе профессиональной деятельности и привлекать для их решения физико-математический аппарат и современные компьютерные технологии

РООПК-4.1 Знает принципы построения технического задания

РООПК-4.2 Умеет использовать нормативные и справочные данные при разработке проектно-конструкторской документации; оформлять проектно-конструкторскую документацию в соответствии со стандартами

РООПК-7.1 Знает оценки эффективности результатов профессиональной деятельности

РООПК-7.2 Умеет выбирать средства и технологии, в том числе с учетом последствий их применения в профессиональной сфере, определять приоритеты профессиональной деятельности и способы ее совершенствования

РООПК-8.1 Знает методы поиска, хранения, обработки, анализа и представления в требуемом формате информации

РООПК-8.2 Умеет решать задачи обработки данных с помощью современных средств автоматизации

РОПК 1.1 Знает основные законы, описывающие функционирование проектируемых объектов.

РОПК 1.2 Умеет использовать стандартные пакеты прикладных программ для выполнения математического моделирования.

РОПК 3.1 Знает основы математического моделирования мехатронных и робототехнических систем.

РОПК 3.2 Умеет использовать стандартные пакеты прикладных программ для выполнения математического моделирования.

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

– тесты, проверяющие (РООПК-2.1, РООПК-2.2, РООПК-4.1, РООПК-4.2, РООПК-7.1, РООПК-7.2, РООПК-8.1, РОПК-8.2, РОПК-1.1, РОПК-1.2, РОПК-3.1, РОПК-3.2);

Тема 1. Основные понятия и задачи

- 1) Что такое машинное обучение?
 - a) Процесс автоматического улучшения программного обеспечения.
 - b) Методология создания искусственного интеллекта.
 - c) Набор методов, позволяющих компьютерам учиться на данных и делать прогнозы без явного программирования правил.
 - d) Технология распознавания образов.
- 2) Какие типы задач решает машинное обучение?
 - a) Классификация, регрессия, кластеризация.
 - b) Прогнозирование погоды, анализ текстов, создание музыки.
 - c) Автоматическое управление роботами, обработка изображений, распознавание речи.
 - d) Все вышеперечисленное.
- 3) Что такое переобучение?
 - a) Ситуация, когда модель слишком хорошо работает на тренировочных данных, но плохо обобщает на новых данных.
 - b) Процесс увеличения количества слоев в нейронной сети.
 - c) Ошибка в коде, приводящая к бесконечному циклу.
 - d) Методика оптимизации гиперпараметров.
- 4) Что такое кросс-валидация?
 - a) Метод оценки качества модели путем разделения данных на тренировочный и тестовый наборы.
 - b) Процедура сравнения нескольких моделей между собой.
 - c) Способ уменьшения шума в данных.
 - d) Алгоритм классификации.
- 5) Что такое регуляризация?
 - a) Процедура нормализации данных.
 - b) Метод предотвращения переобучения путем добавления штрафующих членов к функции потерь.
 - c) Техника повышения точности предсказаний.
 - d) Процесс обучения модели на всех доступных данных.

Тема 2. Основы Python

- 1) Какой оператор используется для присваивания значения переменной в Python?
 - a) =
 - b) :=
 - c) <-
 - d) ==
- 2) Как называется блок кода, который выполняется многократно?
 - a) Класс
 - b) Функция
 - c) Цикл
 - d) Модуль
- 3) Какая функция используется для вывода текста на экран в Python?
 - a) echo()

- b) `output()`
 - c) `display()`
 - d) `print()`
- 4) Как создать список в Python?
- a) `{1, 2, 3}`
 - b) `list(1, 2, 3)`
 - c) `list = [1, 2, 3]`
 - d) `(1, 2, 3)`
- 5) Как вызвать функцию в Python?
- a) `call func()`
 - b) `func()`
 - c) `run(func())`
 - d) `execute(func())`

Тема 3. Работа с данными

- 1) Что такое датафрейм в Pandas?
- a) Таблица данных, аналогичная электронной таблице Excel.
 - b) Структура данных для хранения временных рядов.
 - c) Инструмент для визуализации данных.
 - d) Библиотека для работы с численными данными.
- 2) Как загрузить данные из CSV-файла в Pandas?
- a) `load_data('filename.csv')`
 - b) `import_csv('filename.csv')`
 - c) `pd.read_csv('filename.csv')`
 - d) `data = open('filename.csv').read()`
- 3) Как выбрать столбец из датафрейма в Pandas?
- a) `df['column_name']`
 - b) `select_column(df, 'column_name')`
 - c) `df.column_name`
 - d) `get_column(df, 'column_name')`
- 4) Как удалить строки с пропущенными значениями в Pandas?
- a) `clean_data(df)`
 - b) `remove_missing_values(df)`
 - c) `df.fillna()`
 - d) `df.dropna()`
- 5) Как сгруппировать данные по определенному столбцу в Pandas?
- a) `aggregate(df, 'column_name')`
 - b) `group_data(df, 'column_name')`
 - c) `df.groupby('column_name')`
 - d) `split_by_column(df, 'column_name')`

Тема 4. Подготовка данных и обучение

- 1) Что такое признаки (features) в машинном обучении?
- a) Выходные данные, полученные после обучения модели.
 - b) Входные данные, используемые моделью для обучения и предсказаний.
 - c) Параметры модели, которые необходимо настроить.
 - d) Критерии оценки качества модели.

- 2) Что такое целевая переменная (target variable) в машинном обучении?
 - a) Переменная, которую модель пытается предсказать.
 - b) Переменная, которая описывает процесс обучения модели.
 - c) Переменная, определяющая точность модели.
 - d) Переменная, содержащая информацию о признаках.
- 3) Что такое разбиение данных на тренировочные и тестовые наборы?
 - a) Процедура очистки данных от шумов и выбросов.
 - b) Метод улучшения качества модели путем добавления дополнительных данных.
 - c) Процесс разделения данных на две части: одну для обучения модели, другую для оценки ее производительности.
 - d) Алгоритм подбора оптимальных параметров модели.
- 4) Что такое нормализация данных?
 - a) Удаление дубликатов и пропусков из данных.
 - b) Преобразование данных таким образом, чтобы они находились в определенном диапазоне значений (обычно от 0 до 1).
 - c) Преобразование категориальных признаков в численные.
 - d) Метод оптимизации гиперпараметров.
- 5) Что такое стандартизация данных?
 - a) Процесс преобразования данных в стандартный формат.
 - b) Приведение данных к нулевому среднему и единичной дисперсии.
 - c) Метод снижения размерности данных.
 - d) Процедура удаления выбросов из данных.

Тема 5. Оценка качества моделей машинного обучения.

- 1) Что такое точность (accuracy) в машинном обучении?
 - a) Скорость обучения модели.
 - b) Мера ошибки предсказания.
 - c) Доля правильно классифицированных примеров среди всех примеров.
 - d) Степень сложности модели.
- 2) Что такое полнота (recall) в машинном обучении?
 - a) Время, затраченное на обучение модели.
 - b) Доля отрицательных примеров, неправильно классифицированных как положительные.
 - c) Мера разброса данных.
 - d) Доля положительных примеров, правильно классифицированных как положительные.
- 3) Что такое специфичность (specificity) в машинном обучении?
 - a) Мера ошибки предсказания.
 - b) Доля положительных примеров, правильно классифицированных как положительные.
 - c) Доля отрицательных примеров, правильно классифицированных как отрицательные.
 - d) Скорость обучения модели.
- 4) Что такое кривая ROC (Receiver Operating Characteristic)?
 - a) Кривая, отображающая изменение сложности модели в зависимости от объема данных.
 - b) Кривая, отображающая изменение точности модели в зависимости от времени обучения.
 - c) График, показывающий зависимость ошибки предсказания от числа признаков.

- d) График, показывающий соотношение истинноположительных и ложноположительных результатов при разных порогах классификации.
- 5) Что такое AUC (Area Under the Curve)?
 - a) Площадь под кривой ROC, которая оценивает общую производительность модели.
 - b) Объем памяти, занимаемый моделью.
 - c) Время, необходимое для обучения модели.
 - d) Количество итераций, выполненных моделью.

Тема 6. Нейронные сети.

- 1) Что такое нейронная сеть?
 - a) Метод кластеризации данных.
 - b) Алгоритм классификации, основанный на дереве решений.
 - c) Модель машинного обучения, состоящая из взаимосвязанных узлов, имитирующих структуру мозга.
 - d) Процедура нормализации данных.
- 2) Что такое перцептрон?
 - a) Сложная архитектура нейронной сети, используемая для обработки изображений.
 - b) Простейшая форма нейронной сети, способная выполнять линейную классификацию.
 - c) Алгоритм регрессии.
 - d) Метод понижения размерности данных.
- 3) Что такое скрытый слой в нейронной сети?
 - a) Входной слой нейронов, принимающий данные.
 - b) Последний слой нейронов, выдающий окончательный результат.
 - c) Промежуточный слой нейронов, выполняющий преобразования входных данных.
 - d) Слой нейронов, отвечающий за обратное распространение ошибки.
- 4) Что такое активационная функция в нейронной сети?
 - a) Функция, применяемая к сумме взвешенных входов нейрона для получения выхода.
 - b) Функция, определяющая порядок активации нейронов.
 - c) Функция, рассчитывающая ошибку предсказания.
 - d) Функция, нормализующая данные перед подачей в нейронную сеть.
- 5) Что такое сигмоидная функция?
 - a) Активационная функция, преобразующая входные значения в диапазон от 0 до 1.
 - b) Функция, используемая для расчета ошибки предсказания.
 - c) Функция, применяющаяся для нормализации данных.
 - d) Функция, определяющая порядок активации нейронов.

Тема 7. Фреймворки для работы с моделями машинного обучения.

- 1) Что такое TensorFlow?
 - a) Система управления базами данных.
 - b) Язык программирования для работы с данными.
 - c) Открытая библиотека для машинного обучения, созданная Google.
 - d) Инструмент для визуализации данных.
- 2) Что такое Keras?

- a) API для работы с нейронными сетями, работающий поверх TensorFlow.
 - b) Низкоуровневая библиотека для работы с графиками вычислений.
 - c) Инструмент для анализа данных.
 - d) Система управления версиями.
- 3) Что такое PyTorch?
- a) Инструмент для визуализации данных.
 - b) Библиотека для работы с численными данными.
 - c) Система управления базами данных.
 - d) Фреймворк для машинного обучения, разработанный Facebook AI Research.
- 4) Что такое Scikit-learn?
- a) Инструмент для визуализации данных.
 - b) Инструмент для анализа данных.
 - c) Система управления версиями.
 - d) Библиотека для машинного обучения на Python, предоставляющая готовые реализации многих алгоритмов.
- 5) Что такое MXNet?
- a) Фреймворк, поддерживаемый Apache.
 - b) Библиотека для работы с численными данными.
 - c) Система управления базами данных.
 - d) Инструмент для визуализации данных.

Критерии оценивания: тест считается пройденным, если обучающий ответил правильно как минимум на 4 вопроса.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Промежуточная аттестация осуществляется в форме зачета в письменном виде по билетам. Билет состоит из трех частей.

Первая часть представляет собой тест, проверяющих РООПК-2.1, РООПК-7.1, РООПК-7.2, РООПК-8.1. Ответы на вопросы даются из списка предложенных.

Вторая часть содержит один вопрос, проверяющий РООПК-2.1, РООПК-7.1, РООПК-7.2, РООПК-8.1. Ответ на вопрос дается в развернутой форме.

Третья часть представляет собой результаты практических работ по курсу, выполненных студентом в соответствии с заданием РООПК-2.2, РООПК-4.1, РООПК-4.2, РООПК-8.2, РООПК-1.1, РООПК-1.2, РООПК-3.1, РООПК-3.2).

Тест

1. Что такое машинное обучение?
 - a) Процесс создания компьютерных программ, способных обучаться на данных
 - б) Методика разработки программного обеспечения
 - в) Алгоритм для анализа больших данных
 - г) Способ автоматизации задач программирования
 - д) Набор инструментов для анализа текстов
2. Какие типы задач решаются методами машинного обучения?
 - a) Классификация
 - б) Регрессия
 - в) Кластеризация
 - г) Анализ временных рядов
 - д) Оптимизация
3. Какая из этих моделей относится к методам глубокого обучения?

- а) Линейная регрессия
 - б) K-ближайших соседей
 - в) Нейронная сеть
 - г) Логистическая регрессия
 - д) Дерево решений
4. Что означает термин "обучение с учителем"?
- а) Обучение модели на размеченных данных
 - б) Обучение модели без использования учителя
 - в) Автоматическое обучение модели на основе алгоритмов
 - г) Обучение модели на данных, собранных вручную
 - д) Обучение модели на реальных примерах
5. Что такое переобучение?
- а) Ситуация, когда модель слишком хорошо подходит под тренировочные данные, но плохо обобщает новые данные
 - б) Процесс увеличения количества слоев в нейронной сети
 - в) Ошибка при обучении модели, вызванная недостатком данных
 - г) Процедура оптимизации гиперпараметров модели
 - д) Проблема, связанная с отсутствием достаточного количества вычислительных ресурсов
6. Что такое кросс-валидация?
- а) Метод оценки качества модели путем разделения данных на несколько частей и попеременного тестирования каждой части
 - б) Процедура улучшения точности модели за счет добавления новых данных
 - в) Метод поиска оптимальных параметров модели
 - г) Техника уменьшения размерности данных
 - д) Процесс нормализации данных перед обучением модели
7. Какой алгоритм используется для решения задачи классификации?
- а) Линейная регрессия
 - б) Логистическая регрессия
 - в) SVM (Support Vector Machine)
 - г) Деревья решений
 - д) Все вышеперечисленные
8. Что такое гиперпараметры в контексте машинного обучения?
- а) Параметры модели, которые задаются пользователем перед началом обучения
 - б) Параметры модели, которые автоматически настраиваются в процессе обучения
 - в) Параметры, определяющие сложность модели
 - г) Параметры, влияющие на скорость обучения модели
 - д) Все вышеперечисленное
9. Что такое градиентный спуск?
- а) Алгоритм минимизации функции ошибки
 - б) Метод оптимизации параметров модели
 - в) Процесс обновления весов нейронной сети
 - г) Алгоритм кластеризации данных
 - д) Метод обработки изображений
10. Что такое фича-инжиниринг?
- а) Создание новых признаков на основе исходных данных
 - б) Процесс удаления ненужных признаков из набора данных
 - в) Преобразование данных в формат, удобный для машинного обучения
 - г) Оценка важности признаков для модели
 - д) Все вышеперечисленное
11. Что такое регуляризация?
- а) Метод предотвращения переобучения модели

- б) Процедура нормализации данных
 - в) Алгоритм поиска оптимальной структуры модели
 - г) Метод уменьшения шума в данных
 - д) Процедура очистки данных от выбросов
12. Что такое сверточные нейронные сети (CNN)?
- а) Тип нейронных сетей, специально разработанный для работы с изображениями
 - б) Архитектура нейронных сетей, используемая для обработки текста
 - в) Модель, применяемая для анализа временных рядов
 - г) Тип нейронных сетей, предназначенный для решения задач классификации
 - д) Алгоритм кластеризации данных
13. Что такое рекуррентные нейронные сети (RNN)?
- а) Тип нейронных сетей, способный обрабатывать последовательности данных
 - б) Модель, предназначенная для работы с временными рядами
 - в) Архитектура нейронных сетей, используемая для обработки изображений
 - г) Алгоритм для классификации текстовых данных
 - д) Метод кластеризации данных
14. Что такое ансамбль методов?
- а) Комбинированный подход, использующий несколько моделей для повышения точности предсказаний
 - б) Метод обучения модели на нескольких наборах данных одновременно
 - в) Подход к обучению модели, основанный на использовании различных алгоритмов
 - г) Процедура выбора наилучшей модели из множества предложенных
 - д) Методы визуализации результатов моделирования
15. Что такое точность модели в контексте машинного обучения?
- а) Доля правильно классифицированных примеров среди всех примеров
 - б) Скорость обучения модели
 - в) Время, необходимое для выполнения предсказания
 - г) Степень соответствия модели реальным данным
 - д) Количество ошибок, допущенных моделью
16. Что такое точность (ассигасу) в контексте оценки качества модели?
- а) Максимальная ошибка предсказания
 - б) Процент неправильно классифицированных примеров
 - в) Среднее значение ошибки предсказания
 - г) Процент правильно классифицированных примеров относительно общего числа примеров
 - д) Минимальная ошибка предсказания
17. Что такое F1-мера?
- а) Показатель, который измеряет процент правильно классифицированных отрицательных примеров
 - б) Показатель, который измеряет процент правильно классифицированных положительных примеров
 - в) Средняя взвешенная гармония между точностью и полнотой
 - г) Мера, которая оценивает качество модели на основе максимального значения ошибки
 - д) Мера, которая оценивает качество модели на основе среднего значения ошибки
18. Что такое ROC-кривая?
- а) График, показывающий зависимость точности модели от количества слоев в нейронной сети
 - б) Кривая, отображающая зависимость точности модели от объема данных
 - в) График, показывающий изменение точности модели со временем
 - г) Кривая, отражающая распределение ошибок модели

- д) График, показывающий соотношение истинноположительных и ложноположительных результатов при разных порогах классификации
19. Что такое AUC (Area Under the Curve)?
- а) Отношение площади над кривой к общей площади графика
 - б) Площадь над ROC-кривой, которая показывает степень переобучения модели
 - в) Отношение площади под кривой к общей площади графика
 - г) Площадь под ROC-кривой, которая является мерой качества бинарной классификации
 - д) Площадь под кривой, которая отражает ошибку модели
20. Что такое полнота (recall)?
- а) Доля неправильно классифицированных примеров среди всех примеров
 - б) Доля правильно классифицированных отрицательных примеров среди всех отрицательных примеров
 - в) Доля правильно классифицированных положительных примеров среди всех положительных примеров
 - г) Доля правильно классифицированных примеров среди всех примеров
 - д) Доля неправильно классифицированных положительных примеров среди всех положительных примеров
21. Что такое точность (precision)?
- а) Доля правильно классифицированных примеров среди всех примеров
 - б) Доля правильно классифицированных отрицательных примеров среди всех предсказанных отрицательных примеров
 - в) Доля неправильно классифицированных примеров среди всех примеров
 - г) Доля правильно классифицированных положительных примеров среди всех предсказанных положительных примеров
 - д) Доля неправильно классифицированных положительных примеров среди всех положительных примеров
22. Что такое матрица неточностей (confusion matrix)?
- а) Матрица, показывающая зависимость точности модели от объема данных
 - б) Таблица, представляющая результаты классификации, где строки соответствуют фактическим классам, а столбцы – предсказанным классам
 - в) Таблица, отображающая распределение ошибок модели
 - г) Матрица, показывающая изменения точности модели со временем
 - д) Таблица, отображающая изменения точности модели в зависимости от количества слоев в нейронной сети
23. Что такое логарифмическая потеря (log loss)?
- а) Функция потерь, которая применяется в задачах регрессии
 - б) Функция потерь, которая измеряет разницу между предсказанными значениями и истинными значениями
 - в) Функция потерь, часто используемая в задачах классификации, которая вычисляет вероятность того, что модель делает правильные прогнозы
 - г) Функция потерь, которая используется для измерения степени переобучения модели
 - д) Функция потерь, которая применима только к линейным моделям
24. Что такое метрика MSE (Mean Squared Error)?
- а) Средняя относительная ошибка, которая представляет собой отношение средней
 - б) Средняя абсолютная ошибка, которая представляет собой среднее значение модуля разности между предсказанными и истинными значениями
 - в) Средняя квадратичная ошибка, которая представляет собой среднее значение квадрата разности между предсказанными и истинными значениями квадратичной ошибки к среднему значению истинных значений

- г) Максимальная квадратичная ошибка, которая представляет собой максимальное значение квадрата разности между предсказанными и истинными значениями
- д) Минимальная квадратичная ошибка, которая представляет собой минимальное значение квадрата разности между предсказанными и истинными значениями

25. Что такое метрику MAE (Mean Absolute Error)?

- а) Средняя квадратичная ошибка, которая представляет собой среднее значение квадрата разности между предсказанными и истинными значениями
- б) Средняя абсолютная ошибка, которая представляет собой среднее значение модуля разности между предсказанными и истинными значениями
- в) Средняя относительная ошибка, которая представляет собой отношение средней абсолютной ошибки к среднему значению истинных значений
- г) Максимальная абсолютная ошибка, которая представляет собой максимальное значение модуля разности между предсказанными и истинными значениями
- д) Минимальная абсолютная ошибка, которая представляет собой минимальное значение модуля разности между предсказанными и истинными значениями

Примеры теоретических вопросов:

- 1) Что означает термин "машинное обучение"?
В ответе дать определение термина «машинное обучение» и описать его отличия от традиционных методов анализа данных.
- 2) Чем отличается подход к решению задач в машинном обучении от подхода в классической статистике?
Перечислить ключевые отличия между подходами: сравнение методов, целей, инструментов и результатов. Сделать акцент на уникальных аспектах машинного обучения, таких как адаптивность к большим объемам данных, использование алгоритмов для автоматизации процессов, и способность решать задачи, недоступные традиционным методам статистики.
- 3) Какие основные этапы процесса машинного обучения можно выделить? Какие шаги выполняются на каждом этапе?
При ответе привести подробное описание таких шагов как: выделение задач, определение методов, выбор моделей, настройка параметров, проверка гипотез и оценка эффективности.
- 4) Что такое гиперпараметр в контексте машинного обучения?
При ответе нужно привести примеры гиперпараметров и объяснить зачем важна их настройка.
- 5) Что такое регуляризация в машинном обучении? Почему она необходима и какие подходы к ее применению?
При ответе нужно дать определение, привести примеры методов регуляризации и показать каким образом она может влиять на результат работы модели.
- 6) Что такое переобучение? В каких случаях оно возникает и как его избежать?
При ответе дать определение явления, описать его проявления. Раскрыть причины переобучения и рассказать о способах его избежания.
- 7) Что такое кросс-валидация? Каким образом она помогает оценить качество модели? При ответе дать определение и подробно описать варианты реализации метода с указанием недостатков и описанием способов оценки качества модели.
- 8) Что такое точность (ассигасу) в контексте машинного обучения? Как связаны точность и другие показатели качества модели?
При ответе дать определение, сравнить с другими метриками и описать процесс интерпретации результатов.
- 9) Что такое полнота (recall)? В чем состоит различие между полнотой и точностью?
При ответе описать способ расчета показателей, рассказать почему эти две метрики рассматриваются отдельно.

- 10) Что такое кривая ROC (Receiver Operating Characteristic)? Как интерпретируется площадь под кривой ROC?
Описать способ построения, интерпретации и назначение кривой.
- 11) Что такое метрика F1-score? Как она рассчитывается и как связана с другими показателями качества модели?
При ответе объяснить составляющие компоненты и предоставить формулу расчета. Связать метрику с точностью и полнотой, привести примеры задач, для которых используется метрика.
- 12) Что такое среднеквадратичная ошибка (MSE)? В чем заключается ее смысл и как она применяется в оценке качества модели?
При ответе раскрыть значимость метрики и указать на ограничения.
- 13) Что такое логарифмическая потеря (log loss)? Как она используется для оценки вероятности событий?
При ответе дать определение, раскрыть способ применения и ограничения.
- 14) Что такое сверточная нейронная сеть (CNN)? Для каких типов задач она чаще всего используется?
При ответе дать определение CNN, сделать акцент на преимуществах в обработке пространственных данных и ограничение по вычислительным ресурсам.
- 15) Что такое рекуррентная нейронная сеть (RNN)?
В чем ее особенность и где она находит применение? При ответе описать назначение и структуру модели, описать отличия ее архитектуры от других типов нейронных сетей.
- 16) Что такое перцептрон? Как он связан с современными нейронными сетями?
При ответе рассказать о происхождении перцептрона и как он повлиял на развитие нейросетевых моделей. Рассказать об областях применения перцептронов.
- 17) Что такое обратный перенос ошибки (backpropagation)? Как этот метод используется для обучения нейронных сетей?
В ответе нужно рассказать о механизме обновления весов нейронной сети и описать способы интеграции в нейросетевые модели. Описать альтернативные подходы к обучению нейронных сетей: градиентация и стохастическая аппроксимация.

Пример заданий, выполняемых в рамках практических работ:

1. При помощи библиотеки Scikit-learn создайте модели машинного обучения (дефолтные параметры):
 - Random Forest;
 - Support Vector Machine;
 - Decision Tree;
 - Stochastic Gradient Boosting;
 - k-nearest Neighbors;
 - Gradient Boosting.
2. Подключите и создайте модель машинного обучения Catboost с дефолтными параметрами.
3. При помощи библиотеки Scikit-learn разделите каждый из сформированных ранее наборов данных на тестовые и тренировочные.
4. Примените созданные модели машинного обучения к тестовому и тренировочному наборам. Для каждой из моделей выведите ее конечную точность.
5. Реализуйте k-fold механизм кросс-валидации для каждой из созданных моделей машинного обучения.

6. По результатам работы для каждого из наборов отберите 3 модели машинного обучения, показавшие наилучший результат классификации. Выбор каждой модели обоснуйте письменно в виде комментариев.
7. Произведите поиск значений параметров для отобранных моделей на тестовом и тренировочном наборах, которые позволят повысить их точность. Нужно произвести работу для каждого из наборов и для подобранных для них моделей. В виде комментариев приведите набор наилучших параметров для моделей с выводом об их точности.
8. Проведите поиск наилучших параметров моделей с использованием механизма кросс-валидации.
9. Для 3-х моделей, показавших наилучшие значения по метрике accuracy рассчитайте значения метрик precision, recall, f1-score. Какая из метрик больше подойдет для оценки точности работы моделей?
10. Постройте кривую ROC-AUC для отобранных моделей машинного обучения и в комментариях приведите описание полученных результатов.

Критерии оценивания:

Результат определяется оценками «Зачтено» и «Незачтено».

Оценка «Зачтено» выставляется, если даны правильные ответы на 13 вопросов теста, дан верный развернутый ответ на полученный вопрос и выполнены все практические работы по курсу. В противном случае выставляется оценка «Незачтено».

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Тест

1. Что такое машинное обучение?
 - а) Процесс создания компьютерных программ, способных обучаться на данных
 - б) Методика разработки программного обеспечения
 - в) Алгоритм для анализа больших данных
 - г) Способ автоматизации задач программирования
 - д) Набор инструментов для анализа текстов
2. Какие типы задач решаются методами машинного обучения?
 - а) Классификация
 - б) Регрессия
 - в) Кластеризация
 - г) Анализ временных рядов
 - д) Оптимизация
3. Какая из этих моделей относится к методам глубокого обучения?
 - а) Линейная регрессия
 - б) K-ближайших соседей
 - в) Нейронная сеть
 - г) Логистическая регрессия
 - д) Дерево решений
4. Что означает термин "обучение с учителем"?
 - а) Обучение модели на размеченных данных
 - б) Обучение модели без использования учителя
 - в) Автоматическое обучение модели на основе алгоритмов
 - г) Обучение модели на данных, собранных вручную
 - д) Обучение модели на реальных примерах

5. Что такое переобучение?
- а) Ситуация, когда модель слишком хорошо подходит под тренировочные данные, но плохо обобщает новые данные
 - б) Процесс увеличения количества слоев в нейронной сети
 - в) Ошибка при обучении модели, вызванная недостатком данных
 - г) Процедура оптимизации гиперпараметров модели
 - д) Проблема, связанная с отсутствием достаточного количества вычислительных ресурсов
6. Что такое кросс-валидация?
- а) Метод оценки качества модели путем разделения данных на несколько частей и попеременного тестирования каждой части
 - б) Процедура улучшения точности модели за счет добавления новых данных
 - в) Метод поиска оптимальных параметров модели
 - г) Техника уменьшения размерности данных
 - д) Процесс нормализации данных перед обучением модели
7. Какой алгоритм используется для решения задачи классификации?
- а) Линейная регрессия
 - б) Логистическая регрессия
 - в) SVM (Support Vector Machine)
 - г) Деревья решений
 - д) Все вышеперечисленные
8. Что такое гиперпараметры в контексте машинного обучения?
- а) Параметры модели, которые задаются пользователем перед началом обучения
 - б) Параметры модели, которые автоматически настраиваются в процессе обучения
 - в) Параметры, определяющие сложность модели
 - г) Параметры, влияющие на скорость обучения модели
 - д) Все вышеперечисленное
9. Что такое градиентный спуск?
- а) Алгоритм минимизации функции ошибки
 - б) Метод оптимизации параметров модели
 - в) Процесс обновления весов нейронной сети
 - г) Алгоритм кластеризации данных
 - д) Метод обработки изображений
10. Что такое фича-инжиниринг?
- а) Создание новых признаков на основе исходных данных
 - б) Процесс удаления ненужных признаков из набора данных
 - в) Преобразование данных в формат, удобный для машинного обучения
 - г) Оценка важности признаков для модели
 - д) Все вышеперечисленное
11. Что такое регуляризация?
- а) Метод предотвращения переобучения модели
 - б) Процедура нормализации данных
 - в) Алгоритм поиска оптимальной структуры модели
 - г) Метод уменьшения шума в данных
 - д) Процедура очистки данных от выбросов
12. Что такое сверточные нейронные сети (CNN)?
- а) Тип нейронных сетей, специально разработанный для работы с изображениями
 - б) Архитектура нейронных сетей, используемая для обработки текста
 - в) Модель, применяемая для анализа временных рядов
 - г) Тип нейронных сетей, предназначенный для решения задач классификации
 - д) Алгоритм кластеризации данных
13. Что такое рекуррентные нейронные сети (RNN)?

- а) Тип нейронных сетей, способный обрабатывать последовательности данных
 - б) Модель, предназначенная для работы с временными рядами
 - в) Архитектура нейронных сетей, используемая для обработки изображений
 - г) Алгоритм для классификации текстовых данных
 - д) Метод кластеризации данных
14. Что такое ансамбль методов?
- а) Комбинированный подход, использующий несколько моделей для повышения точности предсказаний
 - б) Метод обучения модели на нескольких наборах данных одновременно
 - в) Подход к обучению модели, основанный на использовании различных алгоритмов
 - г) Процедура выбора наилучшей модели из множества предложенных
 - д) Методы визуализации результатов моделирования
15. Что такое точность модели в контексте машинного обучения?
- а) Доля правильно классифицированных примеров среди всех примеров
 - б) Скорость обучения модели
 - в) Время, необходимое для выполнения предсказания
 - г) Степень соответствия модели реальным данным
 - д) Количество ошибок, допущенных моделью
16. Что такое точность (ассигасу) в контексте оценки качества модели?
- а) Максимальная ошибка предсказания
 - б) Процент неправильно классифицированных примеров
 - в) Среднее значение ошибки предсказания
 - г) Процент правильно классифицированных примеров относительно общего числа примеров
 - д) Минимальная ошибка предсказания
17. Что такое F1-мера?
- а) Показатель, который измеряет процент правильно классифицированных отрицательных примеров
 - б) Показатель, который измеряет процент правильно классифицированных положительных примеров
 - в) Средняя взвешенная гармония между точностью и полнотой
 - г) Мера, которая оценивает качество модели на основе максимального значения ошибки
 - д) Мера, которая оценивает качество модели на основе среднего значения ошибки
18. Что такое ROC-кривая?
- а) График, показывающий зависимость точности модели от количества слоев в нейронной сети
 - б) Кривая, отображающая зависимость точности модели от объема данных
 - в) График, показывающий изменение точности модели со временем
 - г) Кривая, отражающая распределение ошибок модели
 - д) График, показывающий соотношение истинноположительных и ложноположительных результатов при разных порогах классификации
19. Что такое AUC (Area Under the Curve)?
- а) Отношение площади над кривой к общей площади графика
 - б) Площадь над ROC-кривой, которая показывает степень переобучения модели
 - в) Отношение площади под кривой к общей площади графика
 - г) Площадь под ROC-кривой, которая является мерой качества бинарной классификации
 - д) Площадь под кривой, которая отражает ошибку модели
20. Что такое полнота (recall)?
- а) Доля неправильно классифицированных примеров среди всех примеров

- б) Доля правильно классифицированных отрицательных примеров среди всех отрицательных примеров
 - в) Доля правильно классифицированных положительных примеров среди всех положительных примеров
 - г) Доля правильно классифицированных примеров среди всех примеров
 - д) Доля неправильно классифицированных положительных примеров среди всех положительных примеров
21. Что такое точность (precision)?
- а) Доля правильно классифицированных примеров среди всех примеров
 - б) Доля правильно классифицированных отрицательных примеров среди всех предсказанных отрицательных примеров
 - в) Доля неправильно классифицированных примеров среди всех примеров
 - г) Доля правильно классифицированных положительных примеров среди всех предсказанных положительных примеров
 - д) Доля неправильно классифицированных положительных примеров среди всех положительных примеров
22. Что такое матрица неточностей (confusion matrix)?
- а) Матрица, показывающая зависимость точности модели от объема данных
 - б) Таблица, представляющая результаты классификации, где строки соответствуют фактическим классам, а столбцы – предсказанным классам
 - в) Таблица, отображающая распределение ошибок модели
 - г) Матрица, показывающая изменения точности модели со временем
 - д) Таблица, отображающая изменения точности модели в зависимости от количества слоев в нейронной сети
23. Что такое логарифмическая потеря (log loss)?
- а) Функция потерь, которая применяется в задачах регрессии
 - б) Функция потерь, которая измеряет разницу между предсказанными значениями и истинными значениями
 - в) Функция потерь, часто используемая в задачах классификации, которая вычисляет вероятность того, что модель делает правильные прогнозы
 - г) Функция потерь, которая используется для измерения степени переобучения модели
 - д) Функция потерь, которая применима только к линейным моделям
24. Что такое метрика MSE (Mean Squared Error)?
- а) Средняя относительная ошибка, которая представляет собой отношение средней
 - б) Средняя абсолютная ошибка, которая представляет собой среднее значение модуля разности между предсказанными и истинными значениями
 - в) Средняя квадратичная ошибка, которая представляет собой среднее значение квадрата разности между предсказанными и истинными значениями квадратичной ошибки к среднему значению истинных значений
 - г) Максимальная квадратичная ошибка, которая представляет собой максимальное значение квадрата разности между предсказанными и истинными значениями
 - д) Минимальная квадратичная ошибка, которая представляет собой минимальное значение квадрата разности между предсказанными и истинными значениями
25. Что такое метрику MAE (Mean Absolute Error)?
- а) Средняя квадратичная ошибка, которая представляет собой среднее значение квадрата разности между предсказанными и истинными значениями
 - б) Средняя абсолютная ошибка, которая представляет собой среднее значение модуля разности между предсказанными и истинными значениями
 - в) Средняя относительная ошибка, которая представляет собой отношение средней абсолютной ошибки к среднему значению истинных значений

- г) Максимальная абсолютная ошибка, которая представляет собой максимальное значение модуля разности между предсказанными и истинными значениями
- д) Минимальная абсолютная ошибка, которая представляет собой минимальное значение модуля разности между предсказанными и истинными значениями

Критерии оценивания: тест считается пройденным, если обучающий ответил правильно как минимум на половину вопросов.

Информация о разработчиках

Жданов Дмитрий Сергеевич, кандидат технических наук, кафедра прикладной газовой динамики ФТФ, доцент