

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ
Декан факультета информационных
технологий и компьютерной безопасности
Гусев П.Ю.
«21» декабря 2021 г.



**РАБОЧАЯ ПРОГРАММА
дисциплины**

«Машинное обучение и анализ данных»

Направление подготовки 09.04.01 Информатика и вычислительная техника

Профиль Искусственный интеллект

Квалификация выпускника магистр

Нормативный период обучения 2 года / 2 года и 5 м.

Форма обучения очная / заочная

Год начала подготовки 2022

Автор программы

/П.Ю. Гусев/

Заведующий кафедрой
Компьютерных
интеллектуальных
технологий проектирования

/М.И. Чижов/

Руководитель ОПОП

/М.И. Чижов/

Воронеж 2021

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

овладение студентами моделями и методами анализа данных и основ машинного обучения в задачах поиска информации, обработки и анализа данных, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных

1.2. Задачи освоения дисциплины

Изучение инструментов анализа данных.

Способы представления данных для анализа.

Математическое обеспечение анализа данных.

Изучение основных моделей и методов машинного обучения и разработки данных

Приобретение навыков адекватно применять указанные модели и методы, а также программные средства, в которых они реализованы

Приобрести навыки (приобрести опыт) анализа реальных данных с помощью изученных методов

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Машинное обучение и анализ данных» относится к дисциплинам обязательной части блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Машинное обучение и анализ данных» направлен на формирование следующих компетенций:

ПК-6 - Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач

ПК-8 - Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов

| Компетенция | Результаты обучения, характеризующие сформированность компетенции |
|-------------|---|
| ПК-6 | Знать актуальные методы обработки и анализа данных, методы алгоритмов машинного обучения в области применения; |
| | Уметь применять методы машинного обучения при решении задач в различных прикладных областях; уметь подготавливать данные для машинного обучения; использовать библиотеки языка Python для построения моделей машинного обучения |
| | Владеть навыками использования библиотек языка Python для построения интеллектуальных систем. |

| | |
|------|--|
| ПК-8 | Знать сквозные цифровые технологии искусственного интеллекта, включая «Компьютерное зрение», «Обработка естественного языка», «Рекомендательные системы и системы поддержки принятия решений», «Распознавание и синтез речи»; современный опыт применения систем искусственного интеллекта |
| | Уметь проводить сравнительный анализ методов и инструментальных средств для решения задач искусственного интеллекта; разрабатывать и оценивать модели искусственного интеллекта; программировать на языках высокого уровня, ориентированных на применение методов и алгоритмов искусственного интеллекта. |
| | Владеть навыками выявления требований заказчика к результатам решения задач искусственного интеллекта; определения возможностей применения методов искусственного интеллекта в предметной области решаемой задачи; использования имеющейся методологической и технологической инфраструктуры анализа и обработки данных. |

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Машинное обучение и анализ данных» составляет 5 з.е.

Распределение трудоемкости дисциплины по видам занятий
очная форма обучения

| Виды учебной работы | Всего часов | Семестры |
|---|-------------|----------|
| | | 2 |
| Аудиторные занятия (всего) | 34 | 34 |
| В том числе: | | |
| Лекции | 18 | 18 |
| Лабораторные работы (ЛР) | 16 | 16 |
| Самостоятельная работа | 110 | 110 |
| Курсовой проект | + | + |
| Часы на контроль | 36 | 36 |
| Виды промежуточной аттестации - экзамен | + | + |
| Общая трудоемкость: | час | 180 |
| | зач.ед. | 5 |

заочная форма обучения

| Виды учебной работы | Всего часов | Семестры |
|---|-------------|----------|
| | | 3 |
| Аудиторные занятия (всего) | 12 | 12 |
| В том числе: | | |
| Лекции | 4 | 4 |
| Лабораторные работы (ЛР) | 8 | 8 |
| Самостоятельная работа | 159 | 159 |
| Курсовой проект | + | + |
| Часы на контроль | 9 | 9 |
| Виды промежуточной аттестации - экзамен | + | + |
| Общая трудоемкость: | час | 180 |
| | зач.ед. | 5 |
| | | 180 |
| | | 5 |

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

| № п/п | Наименование темы | Содержание раздела | Лекц | Лаб. зан. | СРС | Всего, час |
|--------------|---|--|-----------|-----------|------------|------------|
| 1 | Основные понятия и определения. Примеры прикладных задач | Признаки, вектора признаков. Объекты, классы. Классификация. Классификатор. Обучение, виды обучения "с учителем" и "без учителя". | 4 | 4 | 18 | 26 |
| 2 | Линейные классификаторы | Линейная модель классификации, метод стохастического градиента, алгоритм Персептрона. | 4 | 4 | 18 | 26 |
| 3 | Метод опорных векторов | Основы метода опорных векторов. Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Ядра и спрямляющие пространства. | 4 | 2 | 18 | 24 |
| 4 | Методы восстановления регрессии | Метод наименьших квадратов. Непараметрическая регрессия: ядерное сглаживание. Линейная регрессия. Метод главных компонент. | 2 | 2 | 18 | 22 |
| 5 | Искусственные нейронные сети. Выбор признаков и подготовка данных | Проблема полноты. Задача исключающего "или". Вычислительные возможности двухлинейных и трехслойных сетей. Метод обратного распространения ошибки. Влияние выбора набора признаков на результаты классификации. Предварительная обработка данных. Недостающие значения. Выбор признаков на основе проверки гипотез. Выбор подмножества признаков. | 2 | 2 | 18 | 22 |
| 6 | Контекстно-зависимая классификация | Марковские цепи. Алгоритм Витерби. Скрытые марковские модели. Применение в задачах распознавания голоса. | 2 | 2 | 20 | 24 |
| Итого | | | 18 | 16 | 110 | 144 |

заочная форма обучения

| № п/п | Наименование темы | Содержание раздела | Лекц | Лаб. зан. | СРС | Всего, час |
|-------|--|---|------|-----------|-----|------------|
| 1 | Основные понятия и определения. Примеры прикладных задач | Признаки, вектора признаков. Объекты, классы. Классификация. Классификатор. Обучение, виды обучения "с учителем" и "без учителя". | 2 | 2 | 26 | 30 |

| | | | | | | |
|--------------|---|--|----------|----------|------------|------------|
| 2 | Линейные классификаторы | Линейная модель классификации, метод стохастического градиента, алгоритм Персептрона. | 2 | 2 | 26 | 30 |
| 3 | Метод опорных векторов | Основы метода опорных векторов. Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Ядра и спрямляющие пространства. | - | 2 | 26 | 28 |
| 4 | Методы восстановления регрессии | Метод наименьших квадратов. Непараметрическая регрессия: ядерное сглаживание. Линейная регрессия. Метод главных компонент. | - | 2 | 26 | 28 |
| 5 | Искусственные нейронные сети. Выбор признаков и подготовка данных | Проблема полноты. Задача исключяющего "или". Вычислительные возможности двухлинейных и трехслойных сетей. Метод обратного распространения ошибки. Влияние выбора набора признаков на результаты классификации. Предварительная обработка данных. Недостающие значения. Выбор признаков на основе проверки гипотез. Выбор подмножества признаков. | - | - | 28 | 28 |
| 6 | Контекстно-зависимая классификация | Марковские цепи. Алгоритм Витерби. Скрытые марковские модели. Применение в задачах распознавания голоса. | - | - | 27 | 27 |
| Итого | | | 4 | 8 | 159 | 171 |

5.2 Перечень лабораторных работ

Классификация. Виды обучения "с учителем" и "без учителя".

Алгоритмапостройки нейронных сетей.

Метод стохастического градиента.

Алгоритм Персептрона.

Марковских модель в машинном обучении.

Распознавания голоса.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины предусматривает выполнение курсового проекта в 2 семестре для очной формы обучения, в 3 семестре для заочной формы обучения.

Примерная тематика курсового проекта: «Построение алгоритмов управления методами машинного обучения»

Задачи, решаемые при выполнении курсового проекта:

- Использование методов машинного обучения для разработки алгоритмов идентификации состояний сложных объектов.
- Моделирование полученных алгоритмов управления.
- Анализ полученных алгоритмов управления.

Курсовой проект включают в себя графическую часть и расчетно-пояснительную записку.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

| Компетенция | Результаты обучения, характеризующие сформированность компетенции | Критерии оценивания | Аттестован | Не аттестован |
|-------------|--|-------------------------------|---|---|
| ПК-6 | Знать актуальные методы обработки и анализа данных, методы алгоритмов машинного обучения в области применения; | Количество лабораторных работ | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| | Уметь применять методы машинного обучения при решении задач в различных прикладных областях; уметь подготавливать данные для машинного обучения; использовать библиотеки языка Python для построения моделей машинного обучения | Количество лабораторных работ | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| | Владеть навыками использования библиотек языка Python для построения интеллектуальных систем. | Количество лабораторных работ | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| ПК-8 | Знать сквозные цифровые технологии искусственного интеллекта, включая «Компьютерное зрение», «Обработка естественного языка», «Рекомендательные системы и системы поддержки принятия решений», «Распознавание и синтез речи»; современный опыт применения систем искусственного интеллекта | Количество лабораторных работ | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| | Уметь проводить сравнительный анализ методов и инструментальных средств для решения задач искусственного интеллекта; разрабатывать и оценивать модели искусственного интеллекта; программировать на языках высокого уровня, ориентированных на применение методов и алгоритмов искусственного интеллекта. | Количество лабораторных работ | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| | Владеть навыками выявления требований заказчика к результатам решения задач искусственного интеллекта; определения возможностей применения методов искусственного интеллекта в предметной области решаемой задачи; использования имеющейся методологической и технологической инфраструктуры анализа и обработки данных. | Количество лабораторных работ | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 2 семестре для очной формы обучения, 3 семестре для заочной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

| Компетенция | Результаты обучения, характеризующие сформированность компетенции | Критерии оценивания | Отлично | Хорошо | Удовл. | Неудовл. |
|-------------|--|--|--|---|--|--------------------------------------|
| ПК-6 | Знать актуальные методы обработки и анализа данных, методы алгоритмов машинного обучения в области применения; | Тест | Выполнение теста на 90-100% | Выполнение теста на 80-90% | Выполнение теста на 70-80% | В тесте менее 70% правильных ответов |
| | Уметь применять методы машинного обучения при решении задач в различных прикладных областях; уметь подготавливать данные для машинного обучения; использовать библиотеки языка Python для построения моделей машинного обучения | Решение стандартных практических задач | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены |
| | Владеть навыками использования библиотек языка Python для построения интеллектуальных систем. | Решение прикладных задач в конкретной предметной области | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены |
| ПК-8 | Знать сквозные цифровые технологии искусственного интеллекта, включая «Компьютерное зрение», «Обработка естественного языка», «Рекомендательные системы и системы поддержки принятия решений», «Распознавание и синтез речи»; современный опыт применения систем искусственного интеллекта | Тест | Выполнение теста на 90-100% | Выполнение теста на 80-90% | Выполнение теста на 70-80% | В тесте менее 70% правильных ответов |
| | Уметь проводить сравнительный анализ методов и инструментальных | Решение стандартных | Задачи решены в полном | Продемонстрирован верный ход | Продемонстрирован верный ход решения в | Задачи не решены |

| | | | | | | |
|--|--|--|--|---|--|------------------|
| | средств для решения задач искусственного интеллекта; разрабатывать и оценивать модели искусственного интеллекта; программировать на языках высокого уровня, ориентированных на применение методов и алгоритмов искусственного интеллекта. | практических задач | объеме и получены верные ответы | решения всех, но не получен верный ответ во всех задачах | большинстве задач | |
| | Владеть навыками выявления требований заказчика к результатам решения задач искусственного интеллекта; определения возможностей применения методов искусственного интеллекта в предметной области решаемой задачи; использования имеющейся методологической и технологической инфраструктуры анализа и обработки данных. | Решение прикладных задач в конкретной предметной области | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены |

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1. Какое из следующих утверждений является правильным? (Несколько вариантов)

- A. AdaGrad использует первую производную
- B. L-BFGS использует вторую производную
- C. AdaGrad использует вторую производную
- D. L-BFGS использует первую производную

Ответ : AB

2. «Увеличение размера сверточного ядра, безусловно, улучшит производительность сверточной нейронной сети». Верно ли это утверждение?

- A. Правильно
- B. Ошибка

О т в е т : B

3. Предположим, у вас есть 5 ядер свертки в первом слое сверточной нейронной сети. Каждое ядро свертки имеет размер 7×7 , имеет нулевое заполнение и имеет шаг 1. Размер входного изображения этого слоя составляет $224 \times 224 \times 3$. Каков размер выходного сигнала этого слоя?

- A. $217 \times 217 \times 3$

В. 217 x 217 x 8

С. 218 x 218 x 5

Д. 220 x 220 x 7

О т в е т : С

4. Если существует нейронная сеть и функция активации - ReLU, если вместо ReLU используется линейная функция активации, может ли нейронная сеть по-прежнему характеризовать функцию XNOR?

А. Да

Б. Нет

О т в е т : В

5. Во время обучения машинному обучению размер мини-партии предпочтительно равен степени двойки, например 256 или 512. В чем причина этого?

А. Когда мини-пакет является четным, алгоритм градиентного спуска обучается быстрее

В. Для Mini-Batch установлено значение 2, чтобы удовлетворить требования к памяти процессора и графического процессора и облегчить параллельную обработку.

С. Функция потерь нестабильна, когда четные числа не используются

Д. Ничего из перечисленного

О т в е т : В

6. Какой из следующих методов можно использовать для уменьшения переоснащения? (Несколько вариантов)

А. Больше данных об обучении

Б. L1 регуляризация

С. L2 регуляризация

Д. Уменьшить сложность модели

О т в е т : ABCD

7. Что не так со следующим утверждением?

О. Когда целевая функция выпуклая, решение алгоритма градиентного спуска обычно является глобальным оптимальным решением.

В. Уменьшение размера PCA требует расчета ковариационной матрицы

С. Направление вдоль отрицательного градиента должно быть оптимальным направлением

Д. Лагранжа может решать задачи оптимизации с ограничениями

О т в е т : С

8. Какие из следующих форм алгоритм K-Means может не агрегировать?

А. Круговое распределение

Б. Спиральное распределение

С. Ленточное распространение

Д. Выпуклое распределение полигонов

О т в е т : В

9. Какова норма L1 вектора $X = [1, 2, 3, 4, -9, 0]$?

- A. 1
- B. 19
- C. 6
- D. $\sqrt{111}$

О т в е т : В

10. Какое из следующих утверждений о регуляризации L1 и L2 является правильным?

A. Регуляризация L2 может предотвратить переоснащение и улучшить обобщающую способность модели, но L1 не может этого сделать

B. Технология регуляризации L2 также известна как Регуляризация Лассо.

C. Решение, полученное с помощью регуляризации L1, является более разреженным

D. L2 регуляризационное решение более разреженное

О т в е т : С

7.2.2 Примерный перечень заданий для решения стандартных задач

1. Есть N образцов, которые обычно используются для обучения и обычно используются для тестирования. Если значение N увеличивается, как изменяется разрыв между ошибкой обучения и ошибкой теста?

- A. Увеличение
- B. Уменьшение

О т в е т : В

2. Предположим, вы используете функцию активации X в скрытом слое в нейронной сети. При любом входе определенного нейрона вы получите -0,01. Какие из следующих функций активации может иметь X?

- A. ReLU
- B. tanh
- C. Sigmoid
- D. Все вышеперечисленное возможно

О т в е т : В

3. При каких условиях метод ближайшего соседа k-NN работает лучше?

- A. Больше образцов, но менее типично
- B. Небольшой образец, но хорошая типичность
- C. Образцы распределяются по кластерам.
- D. Образцы распределяются по цепочке

4. Какой из следующих методов можно использовать для сокращения возможностей? (Несколько вариантов)

- A. PCA
- B. Линейный дискриминантный анализ LDA
- C. AutoEncoder
- D. Матрица сингулярного разложения SVD
- E. LeastSquares

5. Какой из следующих методов не может напрямую классифицировать текст?

- A. K-Means
- B. Деревья решений
- C. опорная векторная машина
- D. kNN

О т в е т : А

6. Что в модели регрессии оказывает наибольшее влияние на компромисс между недостаточной и чрезмерной подгонкой?

- A. Полиномиальный порядок
- B. При обновлении весов w используется ли инверсия матрицы или градиентный спуск?
- C. Использование постоянных терминов

О т в е т : А

7. Предположим, у вас есть следующие данные: есть только одна переменная для ввода и вывода. Модель линейной регрессии ($y = wx + b$) была использована для подгонки данных. Так что же является среднеквадратичной ошибкой, полученной с помощью перекрестной проверки Leave-One Out?

| X(independent variable) | Y(dependent variable) |
|-------------------------|-----------------------|
| 0 | 2 |
| 2 | 2 |
| 3 | 1 |

- A. 10/27
- B. 39/27
- C. 49/27
- D. 55/27

О т в е т : С

8. Какое из следующих утверждений об оценке максимального правдоподобия (MLE) является правильным (множественный выбор)?

- A. MLE может не существовать
- B. MLE всегда рядом
- C. Если MLE присутствует, его решение не может быть уникальным
- D. Если MLE существует, его решение должно быть уникальным

О т в е т : АС

9. Если мы говорим, что модель «линейной регрессии» идеально подходит для обучающей выборки (ошибка обучающей выборки равна нулю), какое из следующих утверждений верно?

- A. Ошибка тестового образца всегда равна нулю
- B. Ошибка тестового образца не может быть нулевой
- C. Ни один из приведенных выше ответов

О т в е т : С

10. В задаче линейной регрессии мы используем R-Squared для определения степени соответствия. В настоящее время, если функция добавлена, а модель не изменена, каков следующий оператор?

- A. Если R-Squared увеличивается, эта функция имеет смысл
- B. Если R-Squared уменьшается, эта функция не имеет смысла

С. Просто взглянув на одну переменную R-Squared, невозможно определить, значима ли эта функция.

Д. Ничего из перечисленного

О т в е т : С

7.2.3 Примерный перечень заданий для решения прикладных задач

21. Какое из следующих утверждений об остатках в линейном регрессионном анализе является правильным?

А. Среднее значение остатков всегда равно нулю

В. Средства от остатков всегда меньше нуля

С. Средство остатков всегда больше нуля

Д. Ничего из перечисленного

Ответ : А

22. Какое из следующих утверждений о гетероскедастичности является правильным?

А. Линейная регрессия имеет разные условия ошибки

В. Линейная регрессия имеет тот же термин ошибки

С. Член ошибки линейной регрессии равен нулю

Д. Ничего из перечисленного

О т в е т : А

23. Что из следующего отражает сильную корреляцию между X и Y?

А. Коэффициент корреляции составляет 0,9

В. Значение p для нулевой гипотезы $\beta = 0$ составляет 0,0001.

С. Для неверной гипотезы, $\beta = 0$ с t-значением 30

Д. Ничего из перечисленного

О т в е т : А

24. Каким из следующих предположений мы руководствуемся при получении параметров линейной регрессии (множественный выбор)?

А. X и Y имеют линейные отношения (полиномиальные отношения)

В. Модельные ошибки статистически независимы

С. Ошибки обычно следуют нормальному распределению 0 средних значений и фиксированному стандартному отклонению

Д. X неслучайный и не имеет ошибки измерения

О т в е т : ABCD

25. Чтобы наблюдать линейную зависимость между тестом Y и X, X является непрерывной переменной, какая из следующих графиков больше подходит?

А. Разброс сюжета

В. Гистограмма

С. гистограмма

Д. Ничего из перечисленного

О т в е т : А

26. В целом, какой из следующих методов обычно используется для прогнозирования непрерывных независимых переменных?

- А. Линейная регрессия
- Б. Логический обзор
- С. И линейная регрессия и логистическая регрессия
- Д. Ничего из перечисленного

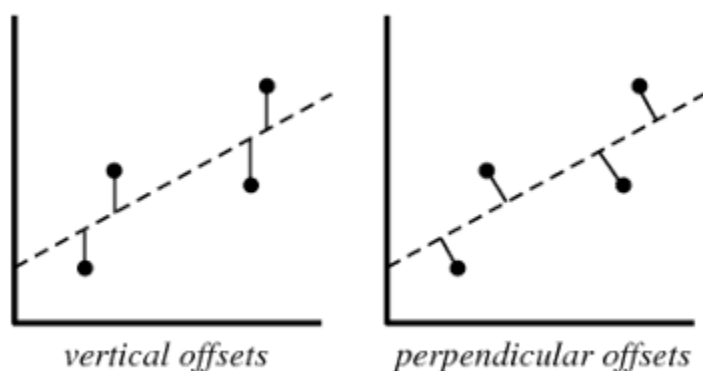
О т в е т : А

27. Коэффициент корреляции между личным здоровьем и возрастом составляет $-1,09$. Какой вывод вы можете сказать доктору на основании этого?

- А. Возраст - хороший показатель здоровья
- Б. Возраст - плохой показатель здоровья
- С. Ничего из перечисленного

О т в е т : С

28. Какие из следующих смещений мы используем в случае подгонки линии наименьших квадратов? На рисунке абсцисса является входом X , а ордината - выходом Y .



- А. Вертикальные смещения
- Б. перпендикулярные смещения
- С. Оба смещения в порядке
- Д. Ничего из перечисленного

О т в е т : А

7.2.4 Примерный перечень вопросов для подготовки к зачету

Не предусмотрено учебным планом

7.2.5 Примерный перечень заданий для подготовки к экзамену

1. Препроцессинг. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding.
2. Кластеризация. kMeans, MeanShift, DBSCAN, Affinity Propagation.
3. Смещение и дисперсия (bias and variance). Понятие средней гипотезы.
4. Ансамблевые методы. Soft and Hard Voting. Bagging. Случайные леса. AdaBoost.
5. Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение.
6. Бустинг деревьев решений.
7. Ошибка внутри и вне выборки. Ошибка обобщения. Неравенство Хёфдинга. Валидация и кросс-валидация.

8. Линейная регрессия. Полиномиальная регрессия. Гребневая регрессия.
9. Размерность Вапника-Червоненкиса. Размерность Вапника-Червоненкиса для перцептрона.
10. Логистическая регрессия. Градиентный спуск.
11. Пороговые условия. Эффективность по Парето. Precision-Recall и ROC кривые. AUC.
12. Ансамблевые методы регрессии. RANSAC. Theil-Sen. Huber.
13. Перцептрон. Перцептрон с карманом.
14. Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра.
15. Гипотезы и дихотомии. Функция роста. Точка поломки. Доказательство полиномиальности функции роста в присутствии точки поломки.
16. Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев. Небрежные решающие деревья.
17. Байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). EM алгоритм.
18. Нейронные сети. Перцептрон Розенблатта. Функции активации. Обратное распространение градиента. Softmax.
19. Стохастическая оптимизация. Hill Climb. Отжиг. Генетический алгоритм.
20. Метрические классификаторы. kNN. WkNN. Отбор эталонов. DROP5. Kdtree

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Экзамен проводится по тест-билетам, каждый из которых содержит 10 вопросов и задачу. Каждый правильный ответ на вопрос в тесте оценивается 1 баллом, задача оценивается в 10 баллов (5 баллов верное решение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20.

1. Оценка «Неудовлетворительно» ставится в случае, если студент набрал менее 6 баллов.
2. Оценка «Удовлетворительно» ставится в случае, если студент набрал от 6 до 10 баллов
3. Оценка «Хорошо» ставится в случае, если студент набрал от 11 до 15 баллов.
4. Оценка «Отлично» ставится, если студент набрал от 16 до 20 баллов.

7.2.7 Паспорт оценочных материалов

| № п/п | Контролируемые разделы (темы) дисциплины | Код контролируемой компетенции | Наименование оценочного средства |
|-------|---|--------------------------------|---|
| 1 | Основные понятия и определения. Примеры прикладных задач | ПК-6, ПК-8 | Тест, защита лабораторных работ, требования к курсовому проекту . |

| | | | |
|---|--|------------|---|
| 2 | Линейные классификаторы | ПК-6, ПК-8 | Тест, защита лабораторных работ, требования к курсовому проекту . |
| 3 | Метод опорных векторов | ПК-6, ПК-8 | Тест, защита лабораторных работ, требования к курсовому проекту . |
| 4 | Методы восстановления регрессии | ПК-6, ПК-8 | Тест, защита лабораторных работ, требования к курсовому проекту . |
| 5 | Искусственные нейронные сети. Выбор признаков и подготовка данных | ПК-6, ПК-8 | Тест, защита лабораторных работ, требования к курсовому проекту . |
| 6 | Контекстно-зависимая классификация | ПК-6, ПК-8 | Тест, защита лабораторных работ, требования к курсовому проекту . |

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Защита курсовой работы, курсового проекта или отчета по всем видам практик осуществляется согласно требованиям, предъявляемым к работе, описанным в методических материалах. Примерное время защиты на одного студента составляет 20 мин.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1 Машинное обучение: новый искусственный интеллект [Текст] : [перевод с английского] / Алпайдин Этем ; Фонд развития пром-сти. - Москва : Издательская группа "Точка", 2017. - 193 с. : ил. - (Завтра это будут знать все). - Библиогр.: с. 185-191. - ISBN 978-5-9908700-8-6

2 Machine Learning: регрессионные методы интеллектуального анализа данных [Электронный ресурс] : Учебное пособие / Л. И. Воронова, В. И. Воронов ; Л. И. Воронова, В. И. Воронов. - Machine Learning: регрессионные методы интеллектуального анализа данных; 2024-02-26. - Москва : Московский технический университет связи и информатики, 2018. - 82 с.

3 Анализ данных [Электронный ресурс] : Учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева ; Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных ; 2024-12-06. - Симферополь : Университет экономики и управления, 2019. - 129 с.

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

Лицензионное ПО:

- Microsoft Word
- IntelliJ PyCharm

Свободное программное обеспечение:

- LibreOffice

Отечественное ПО:

- СУБД Линтер

Ресурсы информационно-телекоммуникационной сети «Интернет»:

- <http://www.edu.ru/>

- Образовательный портал ВГТУ

Информационные справочные системы:

- <http://window.edu.ru>
- <https://wiki.cchgeu.ru/>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Специализированная лекционная аудитория, оснащенная оборудованием для лекционных демонстраций и проекционной аппаратурой.

Учебные лаборатории (г. Воронеж, ул. Плехановская, д. 11):

- 202/2.
- 215/2.

Дисплейный класс, оснащенный компьютерными программами для проведения лабораторного практикума.

Кабинеты, оборудованные проекторами и интерактивными досками.

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Машинное обучение и анализ данных» читаются лекции, проводятся лабораторные работы, выполняется курсовой проект.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Методика выполнения курсового проекта изложена в учебно-методическом пособии. Выполнять этапы курсового проекта должны своевременно и в установленные сроки.

Контроль усвоения материала дисциплины производится проверкой курсового проекта, защитой курсового проекта.

| Вид учебных занятий | Деятельность студента |
|---------------------------------------|--|
| Лекция | Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии. |
| Лабораторная работа | Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания. |
| Самостоятельная работа | Самостоятельная работа студентов способствует глубокому усвоению учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: <ul style="list-style-type: none">- работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций;- выполнение домашних заданий и расчетов;- работа над темами для самостоятельного изучения;- участие в работе студенческих научных конференций, олимпиад;- подготовка к промежуточной аттестации. |
| Подготовка к промежуточной аттестации | Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной |

| | |
|--|---|
| | аттестации. Данные перед экзаменом, экзаменом три дня эффективнее всего использовать для повторения и систематизации материала. |
|--|---|