

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. В.Ф. УТКИНА**

Кафедра «Автоматика и информационные технологии в управлении»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДИСЦИПЛИНЫ
МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Направление 27.03.04

«Управление в технических системах»

ОПОП

«Управление в технических системах»

Квалификация выпускника – бакалавр

Формы обучения – очная

Рязань 2024 г.

Оценочные материалы – это совокупность учебно-методических материалов (контрольных заданий, описаний форм и процедур), предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной профессиональной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретенных компетенций обучающихся целям и требованиям основной профессиональной образовательной программы в ходе проведения текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности общекультурных, общепрофессиональных и профессиональных компетенций, приобретаемых обучающимися в соответствии с этими требованиями.

Контроль знаний обучающихся проводится в форме текущего контроля и промежуточной аттестации.

Текущий контроль успеваемости проводится с целью определения степени усвоения учебного материала, своевременного выявления и устранения недостатков в подготовке обучающихся и принятия необходимых мер по совершенствованию методики преподавания учебной дисциплины (модуля), организации работы обучающихся в ходе учебных занятий и оказания им индивидуальной помощи. К контролю текущей успеваемости относятся проверка знаний, умений и навыков обучающихся: на занятиях; по результатам выполнения лабораторных работ; по результатам выполнения обучающимися индивидуальных заданий.

По итогам курса обучающиеся сдают экзамен. Форма проведения экзамена – устный ответ по утвержденным экзаменационным билетам, сформулированным с учетом содержания учебной дисциплины. В экзаменационный билет включается два теоретических вопроса.

Паспорт фонда оценочных средств по дисциплине (модулю)

№ п/п	Контролируемые разделы (темы) дисциплины (результаты по разделам)	Код контролируемой компетенции (или её части)	Вид, метод, форма оценочного мероприятия
1	2	3	4
1	Раздел 1. Основные понятия теории машинного обучения	ПК-2.1; ПК-2.2	Экзамен
2	Раздел 2. Методы классификации образов	ПК-2.1; ПК-2.2	Экзамен
3	Раздел 3. Методы восстановления регрессии	ПК-2.1; ПК-2.2	Экзамен
4	Раздел 4. Нейросетевые методы классификации и восстановления регрессии	ПК-2.1; ПК-2.2	Экзамен
5	Раздел 5. Кластеризация и обучение без учителя	ПК-2.1; ПК-2.2	Экзамен
6	Раздел 6. Композиция и бустинг алгоритмов	ПК-2.1; ПК-2.2	Экзамен
7	Тема 7. Ранжирование, информационный поиск и коллаборативная фильтрация	ПК-2.1; ПК-2.2	Экзамен

Критерии и шкала оценивания компетенций (результатов)

- 1) Уровень усвоения материала, предусмотренного программой.
- 2) Умение анализировать материал, устанавливать причинно-следственные связи.
- 3) Ответы на вопросы: полнота, аргументированность, убежденность.
- 4) Качество ответа (его общая композиция, логичность, общая эрудиция).
- 5) Использование дополнительной литературы при подготовке ответов.

Успешность усвоения материала оценивается либо оценкой «зачтено», либо - если материал освоен недостаточно - оценкой «не зачтено».

Уровень освоения сформированности знаний, умений и навыков по дисциплине оценивается в форме бальной отметки:

«Отлично» заслуживает студент, обнаруживший всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоивший основную и знакомый с дополнительной литературой, рекомендованной программой. Как правило, оценка «отлично» выставляется студентам,

усвоившим взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявившим творческие способности в понимании, изложении и использовании учебно-программного материала.

«Хорошо» заслуживает студент, обнаруживший полное знание учебно-программного материала, успешно выполняющий предусмотренные в программе задания, усвоивший основную литературу, рекомендованную в программе. Как правило, оценка «хорошо» выставляется студентам, показавшим систематический характер знаний по дисциплине и способным к их самостоятельному пополнению и обновлению в ходе дальнейшей учебной работы и профессиональной деятельности.

«Удовлетворительно» заслуживает студент, обнаруживший знания основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по специальности, справляющийся с выполнением заданий, предусмотренных программой, знакомый с основной литературой, рекомендованной программой. Как правило, оценка «удовлетворительно» выставляется студентам, допустившим погрешности в ответе на экзамене и при выполнении экзаменационных заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя.

«Неудовлетворительно» выставляется студенту, обнаружившему пробелы в знаниях основного учебно-программного материала, допустившему принципиальные ошибки в выполнении предусмотренных программой заданий. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение или приступить к профессиональной деятельности по окончании вуза без дополнительных занятий по соответствующей дисциплине.

Типовые контрольные задания или иные материалы

Вопросы к экзамену по дисциплине (модулю)

ПК-2.1; ПК-2.2

1. Основные понятия теории распознавания образов. Понятие объекта, образа, класса. Особенности распознавания изображений: мешающие факторы.
2. Задача обучения по прецедентам. Понятие обучения, обучающей выборки, контрольной выборки. Понятие машинного обучения и алгоритма распознавания.
3. Признаки и их классификация.
4. Метод обучения. Функционал качества обучения – минимизация эмпирического (среднего) риска. Переобучение.
5. Обобщающая способность. Эмпирические оценки обобщающей способности. Переобучение. Варианты скользящего контроля.
6. Наглядные критерии качества распознавания: матрица ошибок, доля правильных ответов. Точность и полнота. Операционная характеристика бинарного классификатора.

7. Метрические методы классификации. Гипотеза о компактности. Обобщенный метрический классификатор.
8. Метрические методы классификации. Алгоритм ближайшего соседа и его модификации. Метод парзеновского окна.
9. Метод потенциальных функций.
10. Метрические методы классификации. Понятие отступа. Отбор эталонных объектов.
11. Статистические методы классификации – постановка задачи. Оптимальный байесовский классификатор. Задачи эмпирического оценивания.
12. Наивный байесовский классификатор.
13. Статистические методы классификации. Основные подходы к восстановлению плотности распределения. Параметрический и непараметрические подходы к восстановлению плотности.
14. Линейные методы классификации. Обучение классификатора как оптимизационная задача. Непрерывные аппроксимации функции потерь. Градиентный метод численной минимизации функционала эмпирического риска и его вариации.
15. Линейные методы классификации. Проблема переобучения – решение в виде регуляризации. Вероятностная интерпретация регуляризации.
16. Теорема о линейности Байесовского классификатора метод логистической регрессии.
17. Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Постановка и решение задачи нелинейного программирования.
18. Нелинейное обобщение метода опорных векторов: понятия ядра, спрямляющие пространства, конструирование ядер.
19. Постановка задачи восстановления регрессии. Непараметрическая регрессия и метод Надарая-Ватсона. Локально-взвешенное сглаживание, алгоритм LOWESS.
20. Многомерная линейная регрессии, решение с использованием МНК и сингулярного разложения. Регуляризация и отбор признаков.
21. Преобразование признаков по методу главных компонент.
22. Нелинейная регрессия, решение при помощи метода Ньютона-Гаусса. Регрессия с неквадратичными функциями потерь.
23. Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации. Проблема полноты.
24. Алгоритм обратного распространения ошибок. Эвристики: формирование начального приближения, ускорение сходимости, диагональный метод Левенберга-Марквардта.
25. Нейронные сети глубокого обучения. Быстрые методы стохастического градиента. Метод случайных отключений нейронов (Dropout). Функции активации ReLU и PReLU.
26. Свёрточные нейронные сети (CNN). Свёрточный нейрон. Pooling нейрон.
27. Рекуррентные нейронные сети (RNN).
28. Постановка задачи кластеризации. Примеры прикладных задач. Типы

кластерных структур.

29. Оптимизационные постановки задач кластеризации и частичного обучения. Алгоритм k-средних и ЕМ-алгоритм для разделения гауссовской смеси.

30. Графовые алгоритмы кластеризации.

31. Простые эвристические методы частичного обучения: self-training, co-training, co-learning.

32. Композиции классификаторов. Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция. Взвешенное голосование.

33. Алгоритм AdaBoost. Экспоненциальная аппроксимация пороговой функции потерь. Процесс последовательного обучения базовых алгоритмов.

34. Теорема о сходимости бустинга. Обобщающая способность бустинга.

35. Базовые алгоритмы в бустинге. Варианты бустинга. Градиентный бустинг. Стохастический градиентный бустинг.

36. Стохастические методы: бэггинг и метод случайных подпространств.

37. Методы информационного поиска. Меры оценки релевантности документа. Основные подходы к ранжированию.

38. Методы ранжирования RankNet и LambdaRank.

39. Методы коллаборативной фильтрации. Тривиальные рекомендательные системы. Корреляционные алгоритмы.

40. Методы коллаборативной фильтрации. Латентные семантические модели.

Темы практических занятий

ПК-2.1; ПК-2.2

ПР1. Алгоритмы поиска и сортировки в информационно-поисковых системах.

Требуется разработать систему информационного поиска (ИПС) с использованием алгоритмов поиска, сортировки, ранжирования и фильтрации текстов на языке Python. Работа ИПС основана на принципе релевантности (корреляции, совпадения слов) сформированных поисковых образов документов и поисковых запросов по ключевым словам из сформированного по предложенной тематике тезауруса.

ПР2. Разработка системы прогнозирования с использованием линейной регрессии на основе числовых данных на языке Python.

Варианты заданий:

1. Gapminder Dataset: Исследование взаимосвязи между уровнем дохода, продолжительностью жизни, уровнем образования и другими социально-экономическими показателями разных стран.
2. Auto MPG Dataset: Прогнозирование расхода топлива автомобилей на основе характеристик, таких как вес, количество цилиндров, мощность и пробег.
3. Wine Quality Dataset: Оценка качества вина на основе физико-химических характеристик, таких как уровень алкоголя, кислотность, остаточный сахар и др.
4. Iris Dataset: Прогнозирование длины лепестков ирисов на основе ширины лепестков и длины, и ширины чашелистиков.
5. Bike Sharing Dataset: Прогнозирование спроса на аренду велосипедов на основе погодных условий, времени суток, дня недели и сезона.
6. Online News Popularity Dataset: Оценка популярности новостных статей на основе различных метрик, таких как количество слов, количество ссылок на статью, временной интервал и т. д.
7. Air Quality Dataset: Прогнозирование уровня загрязнения воздуха на основе метеорологических данных, времени суток и сезона.
8. Student Performance Dataset: Оценка успеваемости студентов на основе социальных, экономических и академических факторов, таких как пол, возраст, образование родителей и времени, затраченного на учебу.
9. Weather Dataset: Прогнозирование температуры на основе атмосферных условий, таких как влажность, атмосферное давление и скорость ветра.
10. Credit Card Default Dataset: Оценка вероятности невыполнения обязательств по кредитной карте на основе факторов, таких как кредитный лимит, возраст, образование и семейное положение.
11. Energy Efficiency Dataset: Прогнозирование энергоэффективности зданий на основе их характеристик, таких как площадь, высота, ориентация и тип остекления.
12. Forest Fires Dataset: Оценка площади лесных пожаров на основе метеорологических данных и географических координат.
13. Fish Market Dataset: Прогнозирование веса рыбы на основе ее размеров и вида.

14. Black Friday Sales Dataset: Оценка объема продаж на Черную пятницу на основе факторов, таких как возраст, пол, семейное положение и категория товара.

15. Life Expectancy Dataset: Прогнозирование ожидаемой продолжительности жизни на основе таких факторов, как ВВП, уровень образования, наличие питьевой воды и санитарии.

16. Video Game Sales Dataset: Оценка продаж видеоигр на основе жанра, платформы, рейтинга и года выпуска.

17. Employee Attrition Dataset: Прогнозирование текучести кадров на основе факторов, таких как возраст, должность, уровень дохода и удовлетворенность работой.

18. Retail Sales Dataset: Прогнозирование продаж в розничной торговле на основе времени года, праздников, дня недели и погодных условий.

19. Real Estate Price Dataset: Оценка стоимости недвижимости на основе местоположения, площади, числа комнат и других характеристик объекта.

20. E-Commerce Dataset: Прогнозирование объема продаж в электронной коммерции на основе маркетинговых кампаний, сезонности, ассортимента товаров и пользовательских отзывов.

21. Supply Chain Dataset: Оценка времени доставки товаров в зависимости от расстояния, объема груза, маршрута и других факторов.

22. Car Price Dataset: Прогнозирование стоимости автомобилей на основе их характеристик, таких как марка, модель, год выпуска, пробег и состояние.

23. Flight Delay Dataset: Оценка вероятности задержки рейса на основе метеорологических условий, типа самолета, маршрута и времени суток.

24. Telecom Churn Dataset: Прогнозирование оттока клиентов телеком-компаний на основе факторов, таких как срок использования услуг, количество звонков, тип абонента и уровень удовлетворенности.

ПР3. Проектирование и разработка экспертных систем на языке программирования Python для решения различных задач с использованием знаний и опыта экспертов.

Требования к заданию:

1. Выберите область применения для экспертной системы.
2. Определите набор возможных проблем, сценариев или ситуаций, с которыми может столкнуться пользователь, и связанные с ними характеристики или симптомы.
3. Создайте базу знаний, которая будет содержать информацию о соответствии между выбранными сценариями или проблемами и их характеристиками, используя библиотеку LogPy, Pyke или другую подходящую библиотеку.
4. Разработайте пользовательский интерфейс, который позволит вводить данные, характерные для выбранной области (можно использовать текстовый интерфейс или графический).
5. Создайте систему правил для обработки данных пользователя и вывода рекомендаций или решений на основе базы знаний.
6. Проведите тестирование экспертной системы на разных сценариях для проверки корректности выводов и устранения возможных ошибок.
7. Подготовьте отчет, демонстрирующий основные особенности разработанной экспертной системы и результаты работы.

Варианты тем для экспертных систем:

1. Диагностика автомобильных неисправностей.
2. Экспертная система для подбора компьютерного оборудования.
3. Определение подходящего фильма или сериала для просмотра.
4. Экспертная система для подбора книг или научных статей.
5. Определение оптимального плана питания и физической активности.
6. Рекомендации по выбору инвестиционных стратегий на фондовом рынке.
7. Экспертная система для подбора университета и специальности.
8. Подбор косметических средств на основе типа кожи и индивидуальных особенностей.
9. Определение подходящего места для отдыха или путешествий.
10. Экспертная система для анализа рынка недвижимости.
11. Рекомендации по оптимизации энергопотребления в доме или офисе.
12. Подбор музыкальных композиций на основе предпочтений пользователя.

13. Экспертная система для определения нарушений законодательства в договорах и соглашениях.

14. Определение возможных причин технических проблем в сети интернет и предложение решений.

15. Экспертная система для определения подходящих растений для сада на основе климатических условий и типа почвы.

16. Экспертная система для подбора видеоигр на основе предпочтений пользователя и системных требований.

ПК-2.1; ПК-2.2

ПР4. Создание и обучение нейронных сетей для классификации и распознавания различных данных.

Цель: разработать и обучить модель сверточной нейронной сети (CNN) для классификации изображений, используя библиотеку TensorFlow, а затем сравнить ее с моделью, созданной с использованием сервиса Teachable Machine.

1. Выберите или создайте набор данных для задачи классификации изображений. Набор данных должен содержать как минимум 2 класса и не менее 100 изображений для каждого класса.

2. Разделите данные на обучающую, валидационную и тестовую выборки (например, в соотношении 70% на 20% на 10%).

3. Создайте и настройте сверточную нейронную сеть (CNN) с использованием TensorFlow:

- Определите архитектуру сети (количество и типы слоев, функции активации, размеры фильтров и т.д.).
- Настройте гиперпараметры обучения (скорость обучения, размер пакета, количество эпох и т.д.).

4. Обучите нейронную сеть на обучающей выборке, используя валидационную выборку для контроля качества обучения.

5. Оцените качество модели на тестовой выборке, рассчитав метрики, такие как точность, полноту и F1-меру.

6. Визуализируйте результаты классификации, отображая изображения с предсказанными и реальными метками классов.

7. Создайте и обучите модель классификации изображений с использованием сервиса Teachable Machine, загрузив те же обучающие и валидационные данные.

8. Сравните результаты полученной модели с результатами нейронной сети, созданной с использованием TensorFlow.

9. Подготовьте отчет о проделанной работе, включая:

- Описание набора данных, разделения данных и выбранной архитектуры нейронной сети.
- Процесс обучения и настройки гиперпараметров модели.
- Результаты оценки качества модели на тестовой выборке и визуализация результатов классификации.
- Сравнение результатов модели, созданной с использованием TensorFlow, и модели, обученной с помощью сервиса Teachable Machine.

ПК-2.1; ПК-2.2

ПР5. Распознавание объектов на фотографиях с использованием средств OpenCV и Python.

Задание:

1. Выберите набор фотографий, содержащих различные объекты (например, автомобили, людей, животных и т.д.), которые необходимо распознать.

2. Реализуйте программу на Python с использованием библиотеки OpenCV для предобработки фотографий и распознавания объектов на них:

- загрузите изображение и преобразуйте его в оттенки серого;
- примените различные методы обработки изображений (например, размытие, пороговое преобразование, морфологические операции и т.д.), чтобы выделить объекты;
- реализуйте алгоритм распознавания объектов с использованием одного из подходов, предоставляемых библиотекой OpenCV (например, каскады Хаара, HOG+SVM и т.д.);
- нарисуйте ограничивающие прямоугольники вокруг распознанных объектов на изображении и сохраните результат.

3. Протестируйте разработанную программу на выбранном наборе фотографий и оцените качество распознавания объектов.

4. Проведите эксперименты с различными параметрами и настройками алгоритма распознавания объектов для оптимизации качества распознавания.

5. Подготовьте отчет о проделанной работе, включая:

- описание выбранного набора фотографий и объектов для распознавания;

- обзор использованных методов обработки изображений и алгоритмов распознавания объектов;
- описание реализованной программы, ее параметров и настроек;
- результаты тестирования программы на выбранном наборе фотографий и выводы об эффективности распознавания объектов.

ПК-2.1; ПК-2.2

ПР6. Использование алгоритма k-means для кластеризации данных.

Задание:

1. Выберите или создайте набор данных для задачи кластеризации. Набор данных должен содержать не менее 5 признаков.
2. Проведите предварительную обработку данных:
 - Удалите или замените пропущенные значения.
 - Преобразуйте категориальные признаки в числовые.
 - Определите оптимальное количество кластеров с помощью метода «локтя» или других подходов.
3. Создайте и обучите модель k-means, используя библиотеку scikit-learn:
 - Настройте гиперпараметры алгоритма k-means (например, количество кластеров, максимальное количество итераций и т.д.).
 - Обучите модель на обработанных данных.
4. Оцените качество кластеризации:
 - Рассчитайте внутрикластерное расстояние (сумма квадратов расстояний от объектов до их центроидов) и межкластерное расстояние (сумма квадратов расстояний между центроидами).
 - Вычислите силуэтный коэффициент или другие метрики качества кластеризации.
5. Визуализируйте результаты кластеризации:
 - Отобразите данные на двухмерном графике, используя методы снижения размерности, такие как PCA или t-SNE.
 - Подсветите различные кластеры разными цветами и отметьте их центроиды.
6. Проанализируйте полученные кластеры:
 - Исследуйте характеристики объектов внутри каждого кластера.
 - Определите, какие признаки вносят наибольший вклад в формирование кластеров.

- Опишите, как полученные кластеры соответствуют вашим ожиданиям и сформулируйте выводы.
7. Подготовьте отчет о проделанной работе, включая:
- Описание набора данных, предварительной обработки и определения оптимального числа кластеров.
 - Описание процесса обучения модели, используемых гиперпараметров и метрик качества.
 - Визуализацию результатов кластеризации и анализ характеристик объектов внутри кластеров.

Представить описание вклада признаков в формирование кластеров и выводы о соответствии полученных кластеров ожиданиям.

ПК-2.1; ПК-2.2

ПР7. Использование алгоритма деревьев решений для решения задач классификации и анализа результатов.

Задание:

Требуется разработать систему классификации на основе дерева решений с использованием числовых и категориальных признаков на языке Python.

Шаги выполнения:

1. Выберите или создайте набор данных для задачи классификации. Набор данных должен содержать как минимум 2 класса и не менее 5 признаков.
2. Проведите предварительную обработку данных:
 - Удалите или замените пропущенные значения.
 - Преобразуйте категориальные признаки в числовые
 - Разделите данные на обучающую и тестовую выборки (например, в соотношении 80% на 20%).
3. Создайте и обучите модель дерева решений, используя библиотеку scikit-learn:
 - Настройте гиперпараметры дерева решений (например, максимальная глубина, минимальное количество объектов в листе и т.д.).
 - Обучите модель на обучающей выборке.
 - Проведите оценку качества модели:
 - Сделайте предсказания для тестовой выборки.
 - Вычислите метрики качества модели (например, точность, полноту, F1-меру, AUC-ROC и т.д.).
4. Визуализируйте дерево решений и проанализируйте его структуру:
 - Определите наиболее важные признаки.

- Обсудите, как дерево решений разделяет классы на основе признаков.
5. Проведите эксперименты с различными гиперпараметрами и сравните результаты моделей:
- Измените гиперпараметры (например, максимальную глубину, критерий ветвления и т.д.) и сравните качество моделей.
 - Опишите влияние гиперпараметров на качество классификации и структуру дерева.
6. Подготовьте отчет о проделанной работе, включая:
- Описание набора данных, предварительной обработки и разделения данных.
 - Описание процесса обучения модели, используемых гиперпараметров и метрик качества.
 - Визуализацию дерева решений и анализ наиболее важных признаков.
 - Результаты экспериментов с различными гиперпараметрами и их влияние на качество классификации и структуру дерева.
 - Выводы и рекомендации по использованию алгоритма деревьев решений для решения задачи классификации на выбранном наборе данных.

ПК-2.1; ПК-2.2

ПР7. Создание онтологии и выполнения запросов к ней.

Задание:

Требуется разработать онтологию для выбранной предметной области с использованием языка Python.

Шаги выполнения:

1. Выберите одну из предложенных предметных областей.
2. Определите ключевые классы и свойства для вашей предметной области.
3. Установите библиотеку owlready2.
4. Создайте онтологию с использованием библиотеки owlready2 для выбранной предметной области, добавьте в неё классы, свойства и индивиды (экземпляры классов).
5. Создайте несколько SPARQL-запросов для выполнения различных запросов к вашей онтологии (например, поиск объектов по определенным свойствам или классам, агрегация данных и т. д.).
6. Выполните созданные запросы с использованием owlready2 и проанализируйте полученные результаты.
7. Сохраните созданную онтологию в файл формата .owl или .ttl.

8. Подготовьте отчет о проделанной работе, описывающий структуру вашей онтологии, использованные методы и инструменты, а также результаты выполнения запросов к онтологии.

Варианты заданий:

1. Туристические достопримечательности и страны.
2. Виды спортивного оборудования и их назначение.
3. Химические элементы и их свойства.
4. Виды энергии и их источники.
5. Космические объекты и их классификация.
6. Виды финансовых инструментов и их рынки.
7. Виды страхования и их особенности.
8. Климатические зоны и их характеристики.
9. Виды заболеваний и их симптомы.
10. Виды драгоценных камней и их свойства.
11. Методы обучения и их применение.
12. Виды психологических терапий и их подходы.
13. Виды геометрических фигур и их свойства.
14. Виды художественных техник и их применение.
15. Виды экологических систем и их характеристики.
16. Виды транспортных маршрутов и их особенности.
17. Виды сырья и их применение в промышленности.
18. Виды научных исследований и их методология.
19. Здания и архитектурные стили.
20. Профессии и отрасли.
21. Спортивные дисциплины.
22. Блюда и кухни мира.
23. Фильмы и жанры.
24. Музыкальные инструменты.
25. Растения и их классификация.
26. Компьютерная техника.
27. Транспортные средства.
28. Животные и их классификация.

ТЕСТОВЫЕ УПРАЖНЕНИЯ

(ПК-2.1; ПК-2.2)

Упражнение 1.

1.1. Что, из ниже перечисленного, относится к обучающей выборке?

Ответ:

- (1) классификация данных
- + (2) объекты с известными ответами
- (3) алгоритм, решающий функцию

1.2. Объекты состоят из признаков?

Ответ:

- + (1) Да
- (2) Нет

1.3. Что называют данными в машинном обучении?

Ответ:

- + (1) матрицы
- + (2) объекты
- + (3) признаки
- (4) алгоритм
- (5) функция

Упражнение 2.

2.1. Выберите правильный ответ. Задача классификации - это:

Ответ:

- + (1) множество объектов, разделенных на классы
- (2) исследование влияние одного или нескольких признаков на объект
- (3) определение порядка признака согласно рангу

2.2 Выберите правильный ответ. Задача регрессии - это:

Ответ:

- (1) множество объектов, разделенных на классы
- + (2) исследование влияние одного или нескольких признаков на объект
- (3) определение порядка признака согласно рангу

2.3. Выберите правильный ответ. Задача ранжирования - это:

Ответ:

- (1) множество объектов, разделенных на классы
- (2) исследование влияние одного или нескольких признаков на объект
- + (3) определение порядка признака согласно рангу

Упражнение 3.

3.1. Что служит индикатором ошибки для задач классификации?

Ответ:

+ (1) $\varphi(a, x) = [a(x) \neq y^{(*)}(x)]$

(2) $\varphi(a, x) = |a(x) - y^{(*)}(x)|$

(3) $\varphi(a, x) = (a(x) - y^{(*)}(x))^2$

3.2. Как формула подходит для абсолютного значения ошибки для задач регрессии?

Ответ:

+ (1) $\varphi(a, x) = |a(x) - y^{(*)}(x)|$

(2) $\varphi(a, x) = (a(x) - y^{(*)}(x))^2$

(3) $\varphi(a, x) = [a(x) \neq y^{(*)}(x)]$

3.3. Что является квадратичной ошибкой для задачи регрессии?

Ответ:

(1) $\varphi(a, x) = [a(x) \neq y^{(*)}(x)]$

+ (2) $\varphi(a, x) = (a(x) - y^{(*)}(x))^2$

(3) $\varphi(a, x) = |a(x) - y^{(*)}(x)|$

Упражнение 4.

4.1. Эмпирический риск - это средняя потеря на одном объекте.

Ответ:

(1) Да

+ (2) Нет

4.2. Если происходит средняя потеря на всех объектах, то это есть:

Ответ:

(1) переобучение

+ (2) эмпирический риск

(3) оценка релевантности

4.3. Верно ли утверждение? Всякая оптимизация по неполной информации и избыточная сложность параметров приводит в переобучению.

Ответ:

+ (1) Да

(2) Нет

Упражнение 5.

5.1. Выберите верные утверждения.

Ответ:

+ (1) класс - это множество всех объектов с определенным значением.

+ (2) в задачах регрессии допустимым ответом является действительное число или числовой вектор.

+ (3) в задачах ранжирования ответы получают сразу на множестве объектов.

(4) области минимального объёма с достаточно гладкой границей являются основной составляющей задач ранжирования

5.2. Верно ли следующее утверждение? Многие виды задач медицинской диагностики решаются задачами классификации.

Ответ:

+ (1) Да

(2) Нет

5.3. В задачах классификации признаки могут быть строковыми, вещественными, числовыми.

Ответ:

+ (1) Да

(2) Нет

Упражнение 6.

6.1. Какие задачи из ниже перечисленных относятся к задачам классификации?

Ответ:

- + (1) определение наиболее целесообразного способа лечения;
- + (2) определение длительности и исхода заболевания;
- + (3) оценивание кредитоспособности заёмщика;
- (4) задачи поискового вывода

6.2. Какие задачи, из ниже перечисленных, являются задачами ранжирования?

Ответ:

- (1) обнаружение спама
- + (2) задачи поискового вывода;
- + (3) определение наиболее целесообразного способа лечения;

6.3. Какие задачи, из ниже перечисленных, являются задачами прогнозирования?

Ответ:

- + (1) математический прогноз даты сильных землетрясений;
- (2) определение длительности и исхода заболевания;
- (3) обнаружение спама;
- + (4) расчет вероятности летального исхода;
- (5) задачи поискового вывода.

Упражнение 7.

7.1. Какая, из ниже перечисленных задач, является задачей классификации на 4 класса?

Ответ:

(1) $Y = \{0, 1\}^M$

(2) $Y = \{0, 1\}$

(3) $Y = \{-1; +1\}$

+ (4) $Y = \{1, 2, 3, 4\}$

7.2. Какой пример подходит для задачи восстановления регрессии?

Ответ:

(1) $Y = \{0, 1\}^M$

+ (2) $Y = R^m$

(3) $Y = \{-1; +1\}$

(4) $Y = \{1, 2, 3, 4\}$

7.3. Какие, из ниже перечисленных задач, являются задачами классификации?

Ответ:

+ (1) $Y = \{0, 1\}^M$

(2) $Y = R$

+ (3) $Y = \{-1; +1\}$

(4) $Y = R^m$

+ (5) $Y = \{1, 2, 3, 4\}$

Упражнение 8.

8.1. Какой тип экспериментального исследования имеет цель - понимание, на что влияют параметры метода обучения?

Ответ:

- (1) исследование задач ранжирования
- (2) исследование задач классификации
- + (3) исследование на модельных данных

8.2. Какой тип экспериментального исследования имеет цель - либо решение конкретной прикладной задачи, либо выявление «слабых мест»?

Ответ:

- (1) исследование задач ранжирования
- + (2) исследование на реальных данных
- (3) исследование на модельных данных

8.3. Что, из ниже перечисленного, не относится к типу экспериментального исследования?

Ответ:

- + (1) исследование задач ранжирования
- (2) исследование на реальных данных
- (3) исследование на модельных данных