

МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«КОЛЬСКИЙ НАУЧНЫЙ ЦЕНТР РОССИЙСКОЙ АКАДЕМИИ НАУК»
(ФИЦ КНЦ РАН)

УТВЕРЖДАЮ

Начальник управления аспирантуры и магистратуры

ФИЦ КНЦ РАН

к.г.-м.н., доцент И.В. Чикирёв



подпись
29 июня 2020 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

По дисциплине Б1.В.03 Проблемно-ориентированные информационные системы
указывается цикл (раздел) ОП, к которому относится дисциплина, название дисциплины

для направления подготовки (специальности) 09.04.02 Информационные системы и техно-
НОЛОГИИ
код и наименование направления подготовки (специальности)

направленность программы (профиль) Информационные системы предприятий и учре-
ждений
наименование профиля /специализаций/образовательной программы

Квалификация выпускника, уровень подготовки
Магистр

(указывается квалификация (степень) выпускника в соответствии с ФГОС ВО)

Апатиты

2020

Лист согласования

1 Разработчик:

доцент
должность

УАиМ


подпись

Н.А. Тоичкин
И.О. Фамилия

2 Фонд оценочных средств рассмотрен и одобрен на заседании учебно-методической комиссии управления аспирантуры и магистратуры 29 июня 2020 г., протокол № 02.

Председатель УМК УАиМ

29.06.2020
дата


подпись

Л.Д. Кириллова
И.О.Фамилия

Лист переутверждения

Фонд оценочных средств переутвержден на 2021/2022 учебный год без изменений и дополнений.

Председатель УМК УАиМ Л.Д. Кириллова Л.Д. Кириллова

Основание: протокол № 2 от «29» июня 2021 г.

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Лист изменений, вносимых в ФОС по дисциплине «Проблемно-ориентированные информационные системы»

В фонд оценочных средств вносятся следующие изменения и дополнения:

1. _____

2. _____

3. _____

Дополнения и изменения внесены и одобрены на заседании учебно-методической комиссии управления аспирантуры и магистратуры ФИЦ КНЦ РАН

от «____» _____ г., протокол № _____.

Председатель УМК УАиМ _____ Л.Д. Кириллова

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ
АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

Общие сведения

1.	Отдел	Аспирантуры и магистратуры
2.	Направление подготовки	09.04.02 Информационные системы и технологии Профиль – «Информационные системы предприятий и учреждений»
3.	Дисциплина (модуль)	Б1.В.03 Проблемно ориентированные информационные системы
4.	Количество этапов формирования компетенций (ДЕ, разделов, тем и т.д.)	4

Перечень компетенций:

- | |
|---|
| <ul style="list-style-type: none">• Способен проводить экспертизу и оказывать информационно-аналитическую поддержку в решении профессиональных задач в научной деятельности (ПК-1). |
|---|

Критерии и показатели оценивания компетенций на различных этапах их формирования

Этап формирования компетенции (разделы, темы дисциплины)	Формируемая компетенция	Критерии и показатели оценивания компетенций			Формы контроля сформированности компетенций
		Знать:	Уметь:	Владеть:	
1. Большие данные и машинное обучение.	ПК-1	понятие больших данных и их свойства; постановку задачи классификации и регрессии	выполнять постановку задачи машинного обучения	навыками предобработки данных, используя классы библиотеки Pandas	Практическая работа, раздел перечня вопросов к контрольной работе
2. Метрические методы классификации	ПК-1	формализацию задачи; понятие обобщенного метрического классификатора; алгоритмы метрической классификации; метод отбора эталонов, алгоритм STOLP	применять метрические алгоритмы классификации для анализа данных	навыками применения алгоритма k-взвешенных ближайших соседей, используя классы из библиотеки scikit-learn	Практическая работа, раздел перечня вопросов к контрольной работе
3. Логические методы классификации	ПК-1	основные принципы построения логических алгоритмов классификации; критерии информативности: простые критерии, статистический критерий, энтропийный критерий; алгоритм построения дерева ID 3	применять алгоритм ID 3 для классификации данных	навыками работы с алгоритмами классификации на основе деревьев решений, используя классы из библиотеки scikit-learn	Практическая работа, раздел перечня вопросов к контрольной работе
4. Линейные методы классификации	ПК-1	основные принципы построения линейных алгоритмов классификации; алгоритм стохастического градиента; метод SVM	использовать линейные методы для классификации данных	навыками работы с линейными алгоритмами классификации, используя классы из библиотеки scikit-learn	Практическая работа, раздел перечня вопросов к контрольной работе

Критерии и шкалы оценивания

1. Выполнение практической работы

18 баллов выставляется, студент выполнил полностью все задания указанные в лабораторной работе и может аргументировано пояснить ход своего решения.

15 баллов выставляется, если студент выполнил не менее 85 % заданий указанных в лабораторной работе, и может аргументировано пояснить ход своего решения и указать.

11 баллов выставляется, если студент решил не менее 50% заданий указанных в лабораторной работе, и может аргументировано пояснить ход своего решения.

0 баллов выставляется, если студент не может аргументированно пояснить ход своего решения.

В случае если сроки сдачи работ превышены, количество баллов сокращается на 5 баллов.

2. Контрольная работа

Процент правильных ответов	41-60	61-80	81-100
Количество баллов	10	14	18

Оценка	Критерии оценивания ответов контрольной работы
«зачтено»	Обучающийся продемонстрировал знания основных терминов дисциплины, материал изложен последовательно и логично, выводы корректные.
«не зачтено»	Обучающийся не продемонстрировал знания основных терминов дисциплины, материал изложен непоследовательно и нелогично, выводы некорректные.

Пример типового тестового задания (контрольной работы)

1. К какому типу задач машинного обучения относится задача предсказания цены жилья по его характеристикам?

- : Классификация на два класса;
- : Классификация на M непересекающихся классов;
- : Классификация на M пересекающихся классов;
- +: Восстановление регрессии;

2. К какому типу относится признаки «Цвет глаз»?

- : Бинарный;
- : Количественный;
- +: Номинальный (категориальный);
- : Порядковый;

3. К какому типу задач машинного обучения, относится задача в которой необходимо определить независимые группы и их характеристики во всем множестве анализируемых данных?

- : задача классификации;
- : задача регрессии;
- +: задача кластеризации;

4. К какому типу задач машинного обучения, относится задача в которой необходимо определить зависимости между объектами или событиями?

- задача распознавания образов;
- + задача поиска ассоциативных правил;
- задача нормализации;

5. Астроном Витя хочет построить модель, которая сможет разбить известные науке звезды на группы по их характеристикам, чтобы лучше изучить их особенности. К какому типу относится данная задача?

- + Кластеризация
- Классификация
- Ранжирование
- Регрессия

6. Выберите все верные утверждения:

- Модель машинного обучения, по сути, является отображением пространства ответов в пространство объектов
- Функционал ошибки показывает, насколько плохое качество имеют данные, используемые для решения задачи
- + Процесс обучения модели заключается в минимизации функционала ошибки
- + Элементами обучающей выборки являются объекты, характеристики которых являются значениями признаков

7. Можно ли из модели линейной регрессии выбросить свободный коэффициент ω_0 и почему?

-: Нет, потому что без ω_0 модель будет константной

+: Нет, потому что без свободного коэффициента модель гарантированно будет давать нулевой прогноз при нулевых значениях всех признаков, а это ограничивает возможности по подгонке под данные

-: Да, он участвует в модели только для упрощения процедуры обучения, но при этом никак не повышает её силу

-: Да, он участвует в модели исключительно по историческим причинам

8. Рассмотрим признак “Образовательная программа” при анализе данных по студентам университета. Этот признак может принимать три значения: “Экономика”, “Математика”, “Философия”. Воспользуемся one-hot кодированием и заменим этот признак на три бинарных, которые будут соответствовать категориям в том порядке, в котором они перечислены выше. Как будет закодирован признак со значением “Философия”?

-: (0, 1, 0)

-: (1, 0, 0)

+: (0, 0, 1)

9. Предположим, что мы строим модель предсказания роста по возрасту и весу человека. Модель с какими коэффициентами вероятнее всего переобучилась?

-: $0.001 * (\text{возраст}) + 0.5 * (\text{вес})$

+: $1402325.3 * (\text{возраст}) + -1404370.5 * (\text{вес})$ верно

-: $0.1 * (\text{возраст}) + 0.33 * (\text{вес})$

10. Предположим, что мы строим модель предсказания стоимости дома по количеству комнат и средней цене дома в районе. Перед количеством комнат коэффициент равен 1400230, а перед средней ценой дома в районе 0.8. Можно ли утверждать, что количество комнат — более важный признак для качества предсказания, чем средняя цена в районе и почему?

-: Да, так как количество комнат — это признак, который может принимать небольшое количество значений, а значит, каждое значение содержит в себе больше информации

+: Нет, так как коэффициенты несравнимы, поскольку признаки имеют разный масштаб

-: Нет, так как средняя цена дома в районе — это признак с большим разбросом, а именно разброс характеризует ценность признака

-: Да, так как коэффициент перед количеством комнат больше

11. Заполните пропуск: «Для выпуклой функции указывает сторону наискорейшего убывания»

-: Изохрона

-: Линия уровня

+: Антиградиент

-: Градиент

12. Позволяет ли градиентный спуск находить минимум функции?

- : Нет, процесс подбора не сходится к минимуму
- : Да, алгоритм всегда сходится к глобальному минимуму
- +: Да, но только локальный минимум

13. Найдите евклидову норму вектора $x = (1, 1, 1.5, 1)$. Ответ вводите с точкой, с точностью до 2-х знаков после запятой.

+: 2.29

14. В лекции был рассказан алгоритм градиентного спуска, в котором на каждом шаге параметры сдвигаются в сторону антиградиента, за счёт чего уменьшается ошибка. Представьте себе другой алгоритм: на каждом шаге мы выбираем один из параметров и сдвигаем его по нему влево или вправо (выбираем такое направление, чтобы ошибка уменьшалась сильнее). Такой метод называют покоординатным спуском. Чем он хуже?

- +: Покоординатный спуск, скорее всего, потребует большего числа итераций, чем градиентный спуск.
- : Покоординатный спуск требует подсчёта градиентов, а это может быть достаточно затратно по времени.
- : Покоординатный спуск не даёт никаких гарантий того, что ошибка будет уменьшаться с числом шагов.
- +: Покоординатный спуск может быть неэффективным в пространствах большой размерности, поскольку даже чтобы сдвинуться по каждой координате по одному разу, потребуется много шагов.

15. Чем глобальный минимум отличается от локального?

- : В точке глобального минимума функция принимает уникальное значение, которое не достигается больше нигде.
- : В точке глобального минимума функция принимает самое маленькое значение по сравнению с точками, находящимися рядом.
- +: В точке глобального минимума функция принимает такое значение, что меньше него достичь нельзя.

16. Чем стохастический градиентный спуск (SGD) лучше обычного градиентного спуска? Выберите все подходящие ответы.

- +: Один шаг в SGD быстрее, поэтому в целом этот метод может быстрее выдать решение.
- : В SGD гарантируется, что на каждой итерации уменьшается ошибка модели.
- : В SGD гарантируется, что будет найден глобальный минимум.
- : Один шаг в SGD точнее, чем в обычном градиентном спуске, поэтому требуется меньше шагов для получения решения.

17. Вы решаете задачу классификации писем на нормальные и спам, причём примеров спама очень мало (меньше 1% от размера всей выборки). Вы решили рассмотреть модель, которая для любого письма говорит, что оно нормальное. Выберите все верные утверждения про метрики качества для такой модели.

- : Доля верных ответов будет близка к 50%
- +: Полнота будет равна единице
- +: Доля верных ответов будет высокой
- : Точность будет равна единице

18. Что оценивает абсолютное значение отступа?

- : Модуль веса, соответствующего данному объекту
- : Вероятность класса для данного объекта
- +: Расстояние от объекта до разделяющей гиперплоскости
- : Вероятность того, что данный объект является выбросом

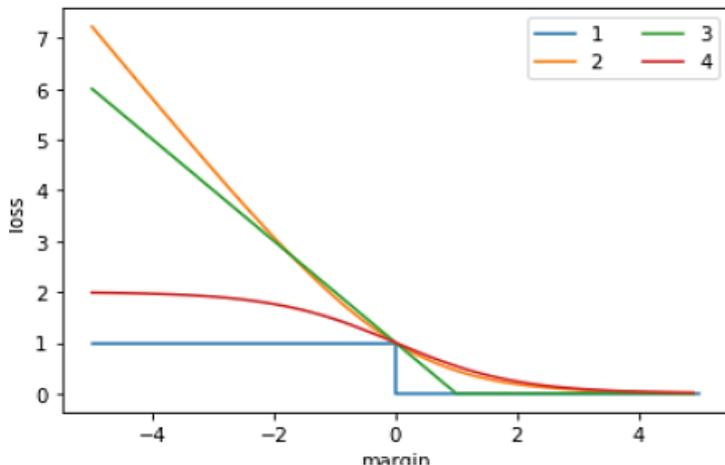
19. Рассмотрим выборку из трех объектов, принадлежащих к классам 0, 1 и 1 соответственно. Оценка принадлежности классу 1 алгоритма классификации для первого объекта равна 0.2, для второго - 0.4, и для третьего - 0.9. Найдите площадь под PR-кривой для данного классификатора на данной выборке.

+: 0.75

20. Выберите верные утверждения про метод опорных векторов для линейно неразделимого случая:

- +: В методе опорных векторов максимизируется расстояние от разделяющей поверхности до ближайшего объекта обучающей выборки
- : В методе опорных векторов минимизируется расстояние от разделяющей поверхности до ближайшего объекта обучающей выборки
- +: В методе опорных векторов на объекте может быть допущена ошибка, но за это в функционале добавляется штраф
- : В методе опорных векторов запрещено допускать ошибки на объектах обучающей выборки

21. Какая аппроксимация функции потерь используется в логистической регрессии? Выберите график этой функции.



- : 1
- +: 2
- : 3
- : 4

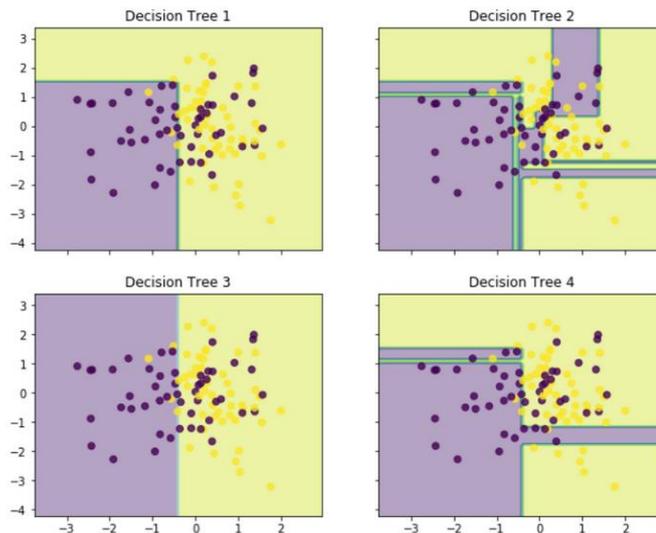
22. Выберите верные утверждения про решающие деревья:

- +: В каждой внутренней вершине дерева проверяется некоторое условие
- : В каждой внутренней вершине дерева выдается некоторый прогноз
- +: Чем глубже дерево, тем более сложная получается модель
- : Чем глубже дерево, тем менее переобученная получается модель
- : С помощью решающего дерева можно идеально решить задачи только с линейно разделимой выборкой

23. Выберите верные утверждения про измерение энтропии при разбиении вершины:

- + : Энтропия вычисляется для распределения классов в вершине
- : Энтропия вычисляется для распределения одного из признаков в вершине
- + : Чем меньше энтропия в вершине, тем больше там объектов одного класса
- : Чем больше энтропия в вершине, тем больше там объектов одного класса
- + : Разумно выбирать такое разбиение, в котором энтропия дочерних вершин как можно меньше
- : Разумно выбирать такое разбиение, в котором энтропия дочерних вершин как можно больше

24. На рисунках изображены разделяющие поверхности для задачи бинарной классификации, соответствующие решающим деревьям разной глубины. Какое из изображений соответствует наиболее глубокому дереву?



- : Decision Tree 1
- + : Decision Tree 2
- : Decision Tree 3
- : Decision Tree 4

25. Используя приведенную таблицу, по какому признаку следует формировать первый узел решающего дерева, если мы хотим предсказать Y? В качестве критерия информативности использовать энтропию, в качестве критериев разделения - индикаторы $[x_j = a]$.

X1	X2	X3	X4	Y
A1	A2	A3	A4	A
B1	A2	B3	A4	A
C1	C2	A3	A4	A
A1	A2	D3	B4	A
C1	B2	C3	A4	B
B1	C2	D3	B4	A
A1	B2	B3	A4	A
C1	C2	C3	B4	B
B1	B2	C3	B4	B
A1	C2	C3	A4	B

- : X1
- : X2
- + : X3
- : X4

Вопросы к зачету по дисциплине

«Проблемно ориентированные информационные системы»

1. Основные понятия – информация, данные, знания. Виды информации. Обработка данных и ее виды. Data Mining. Классификация задач Data Mining.
2. Модели процессов обработки данных. Модель: конечные автоматы.
3. Модели процессов обработки данных. Модель: сети Петри.
4. Задачи обработки данных различных типов. Прикладные области обработки данных. Оцифровка сигналов. Теорема Котельникова.
5. Базы данных. OLTP – системы. Неэффективность OLTP для анализа данных. Определение и свойства хранилищ данных.
6. Физические и виртуальные хранилища данных (ХД). Основные проблемы создания ХД.
7. Витрины данных.
8. Данные в хранилищах данных. ETL процесс.
9. Представление данных в виде гиперкуба. Операции над гиперкубом. Пример. Технология OLAP. Тест FASMI.
10. Многомерное представление данных и многомерный куб. Представление данных в виде гиперкуба. Пример.
11. Основные понятия гиперкубов (OLAP кубов). Структура OLAP куба. Операции над гиперкубом.
12. Архитектура OLAP. Компоненты OLAP. MOLAP, ROLAP, HOLAP.
13. Задача анализа текстов. Этапы анализа. Предобработка текста.
14. Извлечение ключевых понятий из текста.
15. Классификация текстовых документов. Методы классификации текстовых документов.
16. Большие данные. Свойства больших данных.
17. Машинное обучение, формализация задачи машинного обучения.
18. Признаковое описание объекта. Ответы и типы задач машинного обучения. Модель алгоритмов. Метод обучения. Этап обучения и этап применения.
19. Функционалы качества. Сведение задачи обучения к задаче оптимизации.
20. Переобучение и обобщение. Пример переобучения (Рунге). Эмпирические оценки обобщающей способности.
21. Примеры задач машинного обучения: задачи классификации.
22. Примеры задач машинного обучения: задачи регрессии.
23. Примеры задач машинного обучения: задача ранжирования.
24. Эксперименты в машинном обучении: эксперименты на реальных и синтетических данных.
25. Формализация метрической классификации. Обобщенный метрический классификатор.
26. Метод ближайшего соседа.
27. Метод k взвешенных ближайших соседей.
28. Метод парзеновского окна.
29. Метод потенциальных функций.
30. Отбор эталонных объектов. Понятие отступа объекта. Типы объектов в зависимости от отступа.
31. Отбор эталонов, алгоритм STOLP.
32. Логическая закономерность. Основы вопросы построения логических алгоритмов классификации. Виды закономерностей.
33. Критерии информативности: простые критерии, статистический критерий, энтропийный критерий. Схема локального поиска информативных закономерностей.

34. Определение бинарного решающего дерева. Жадный алгоритм построения дерева ID 3.
35. Варианты критериев ветвления в ID 3.
36. Алгоритм ID3: достоинства и недостатки.
37. Обработка пропусков в ID 3, алгоритм обработки пропусков на этапе обучения и этапе классификации.
38. Стратегии редукции решающих деревьев.
39. Небрежные решающие деревья.
40. Бинаризация вещественного признака.