

ПРИЛОЖЕНИЕ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Рязанский государственный радиотехнический университет им. В.Ф. Уткина»

КАФЕДРА ЭЛЕКТРОННЫЕ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

по дисциплине

«Основы систем искусственного интеллекта»

Направление подготовки

02.03.03 – «Математическое обеспечение и администрирование информационных систем»

Направленность (профиль) подготовки

«Программное обеспечение компьютерных технологий и систем искусственного интеллекта»

Уровень подготовки - бакалавриат

Квалификация выпускника – бакалавр

Форма обучения – очная, очно-заочная

Рязань

1 ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов (практических заданий, описаний форм и процедур проверки), предназначенных для оценки качества освоения обучающимися данной дисциплины как части ОПОП.

Цель – оценить соответствие знаний, умений и владений, приобретенных обучающимся в процессе изучения дисциплины, целям и требованиям ОПОП в ходе проведения промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций. Контроль знаний обучающихся проводится в форме промежуточной аттестации. Промежуточная аттестация проводится в форме экзамена. Форма проведения экзамена - тестирование, письменный опрос по теоретическим вопросам и выполнение практического задания.

2 ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ

Сформированность каждой компетенции (или ее части) в рамках освоения данной дисциплины оценивается по трехуровневой шкале:

- 1) пороговый уровень является обязательным для всех обучающихся по завершении освоения дисциплины;
- 2) продвинутый уровень характеризуется превышением минимальных характеристик сформированности компетенций по завершении освоения дисциплины;
- 3) эталонный уровень характеризуется максимально возможной выраженностью компетенций и является важным качественным ориентиром для самосовершенствования.

Уровень освоения компетенций, формируемых дисциплиной:

Описание критериев и шкалы оценивания тестирования:

Шкала оценивания	Критерий
3 балла (эталонный уровень)	уровень усвоения материала, предусмотренного программой: процент верных ответов на тестовые вопросы от 85 до 100%
2 балла (продвинутый уровень)	уровень усвоения материала, предусмотренного программой: процент верных ответов на тестовые вопросы от 70 до 84%
1 балл (пороговый уровень)	уровень усвоения материала, предусмотренного программой: процент верных ответов на тестовые вопросы от 50 до 69%
0 баллов	уровень усвоения материала, предусмотренного программой: процент верных ответов на тестовые вопросы от 0 до 49%

Описание критериев и шкалы оценивания теоретического вопроса:

Шкала оценивания	Критерий
3 балла (эталонный уровень)	выставляется студенту, который дал полный ответ на вопрос, показал глубокие систематизированные знания, смог привести примеры, ответил на дополнительные вопросы преподавателя
2 балла (продвинутый уровень)	выставляется студенту, который дал полный ответ на вопрос, но на некоторые дополнительные вопросы преподавателя ответил только с помощью наводящих вопросов
1 балл (пороговый уровень)	выставляется студенту, который дал неполный ответ на вопрос в билете и смог ответить на дополнительные вопросы только с помощью преподавателя

<i>0 баллов</i>	выставляется студенту, который не смог ответить на вопрос
-----------------	---

Описание критериев и шкалы оценивания практического задания:

<i>Шкала оценивания</i>	<i>Критерий</i>
<i>3 балла (эталонный уровень)</i>	Задача решена верно
<i>2 балла (продвинутый уровень)</i>	Задача решена верно, но имеются неточности в логике решения
<i>1 балл (пороговый уровень)</i>	Задача решена верно, с дополнительными наводящими вопросами преподавателя
<i>0 баллов</i>	Задача не решена

На промежуточную аттестацию выносится тест, два теоретических вопроса и задача. Максимально студент может набрать 12 баллов. Итоговый суммарный балл студента, полученный при прохождении промежуточной аттестации, переводится в традиционную форму по системе «отлично», «хорошо», «удовлетворительно» и «неудовлетворительно».

Оценка «отлично» выставляется студенту, который набрал в сумме 12 баллов (выполнил все задания на эталонном уровне). Обязательным условием является выполнение всех предусмотренных в течение семестра практических заданий.

Оценка «хорошо» выставляется студенту, который набрал в сумме от 8 до 11 баллов при условии выполнения всех заданий на уровне не ниже продвинутого. Обязательным условием является выполнение всех предусмотренных в течение семестра практических заданий.

Оценка «удовлетворительно» выставляется студенту, который набрал в сумме от 4 до 7 баллов при условии выполнения всех заданий на уровне не ниже порогового. Обязательным условием является выполнение всех предусмотренных в течение семестра практических заданий.

Оценка «неудовлетворительно» выставляется студенту, который набрал в сумме менее 4 баллов или не выполнил всех предусмотренных в течение семестра практических заданий.

3 ПАСПОРТ ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ

<i>Контролируемые разделы (темы) дисциплины</i>	<i>Код контролируемой компетенции (или её части)</i>	<i>Вид, метод, форма оценочного мероприятия</i>
Тема 1. Задача классификации в ИИ. Метрическая классификация.	УК-11.1 УК-11.2	Экзамен
Тема 2. Линейная классификация.	ПК-19.1 ПК-19.2	Экзамен
Тема 3. Линейная регрессия.	ПК-19.1 ПК-19.2	Экзамен
Тема 4. Байесовские методы классификации.	ПК-19.1 ПК-19.2	Экзамен
Тема 5. Кластеризация.	ПК-19.1 ПК-19.2 ОПК-6.1 ОПК-6.2	Экзамен

Тема 6. Основы нейронных сетей.	ПК-19.1 ПК-19.2	Экзамен
Тема 7. Принципы обучения нейронных сетей.	ПК-19.1 ПК-19.2	Экзамен
Тема 8. Основные архитектуры нейронных сетей.	ПК-19.1 ПК-19.2	Экзамен

4 ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ

4.1. Промежуточная аттестация в форме экзамена

Код компетенции	Результаты освоения ОПОП Содержание компетенций
УК-11	Способен осуществлять свою трудовую деятельность с учетом определения корректной роли ИИ в различных процессах, критического анализа последствий применения ИИ-технологий, этических принципов

УК-11.1. Определяет ценностные предпосылки, когнитивные искажения, культурно-обусловленные предвзятости в данных, алгоритмах, постановке задач для ИИ Типовые тестовые вопросы

Типовые тестовые вопросы

1. Что такое функция расстояния (метрика) в контексте метрических алгоритмов?

- a) Функция, которая преобразует данные в новый признаковое пространство.
- b) Функция, удовлетворяющая условиям неотрицательности, тождества неразличимых, симметрии и неравенству треугольника.
- c) Функция, которая измеряет сходство объектов, всегда возвращая значение между 0 и 1.
- d) Функция, используемая исключительно для нормализации данных.

Правильный ответ: b)

2. Каков ключевой недостаток наивного подхода к реализации метода k ближайших соседей (k-NN) при работе с большими данными?

- a) Низкая точность прогноза.
- b) Высокая вычислительная сложность классификации нового объекта.
- c) Неспособность работать с категориальными признаками.
- d) Чувствительность к выбору начальных центроидов.

Правильный ответ: b)

3. Какой из перечисленных методов НЕ является методом ускорения поиска ближайших соседей?

- a) KD-дерево.
- b) Алгоритм баллистического поиска.
- c) Локально-чувствительное хеширование (LSH).
- d) Приближенный поиск по метрическим пространствам.

Правильный ответ: b)

4. Зачем нужна нормализация данных перед применением метрического классификатора?

- a) Для увеличения скорости сходимости алгоритма оптимизации.
- b) Для того чтобы признаки с большим разбросом значений не доминировали при расчете расстояния.
- c) Для автоматического отбора наиболее значимых признаков.
- d) Для преобразования нелинейных зависимостей в линейные.

Правильный ответ: b)

5. Что оценивает процедура кросс-валидации?

- a) Скорость работы алгоритма на тестовых данных.
- b) Способность алгоритма к обобщению на ранее невиданных данных.
- c) Вычислительную сложность алгоритма.
- d) Оптимальное количество параметров в модели.

Правильный ответ: b)

6. При использовании метода k-NN, что может произойти, если выбрать значение k слишком маленьким (например, k=1)?

- a) Модель станет излишне сложной и будет шуметь в данных.
- b) Модель станет чрезмерно сглаженной и не сможет уловить сложные зависимости.
- c) Модель станет более устойчивой к выбросам.
- d) Вычислительная сложность алгоритма значительно возрастет.

Правильный ответ: a)

7. Как вариация метода ближайших соседей, известная как "взвешенные ближайшие соседи", модифицирует базовый алгоритм?

- a) Присваивает больший вес соседям, которые находятся дальше от классифицируемого объекта.
- b) Автоматически подбирает оптимальное значение k для каждого объекта.
- c) Присваивает больший вес более близким соседям при принятии решения о классе.
- d) Использует другую метрику расстояния для более точных соседей.

Правильный ответ: c)

8. Какова основная цель предобработки данных перед построением метрической модели?

- a) Увеличить размер обучающей выборки.
- b) Упростить алгоритм классификации.
- c) Повысить качество и надежность модели за счет улучшения исходных данных.
- d) Уменьшить размерность данных до 2 или 3 для визуализации.

Правильный ответ: c)

9. Какой из перечисленных методов НЕ является методом ускорения поиска ближайших соседей?

- a) KD-дерево
- b) Метод опорных векторов (SVM)
- c) Локально-чувствительное хеширование (LSH)
- d) Приближенный поиск ближайших соседей

10. Какой из перечисленных методов предобработки масштабирует данные таким образом, что каждый признак имеет среднее значение 0 и стандартное отклонение 1?

- a) Минимаксная нормализация (например, до [0, 1])
- b) Нормализация по длине вектора
- c) Стандартизация (Z-score нормализация)
- d) Нормализация с помощью логарифмирования

Правильный ответ: c)

Типовые вопросы открытого типа:

1. Дайте строгое математическое определение метрики. Какие четыре условия должна удовлетворять функция, чтобы называться метрикой?
2. В чем заключается ключевое различие между метрическими и параметрическими методами классификации? Приведите примеры тех и других.
3. Опишите алгоритм k ближайших соседей (k-NN) своими словами. Как происходит процесс классификации нового объекта?
4. Какое влияние на модель k-NN оказывает выбор малого ($k=1$) и очень большого значения k ? С какими проблемами мы можем столкнуться в каждом из этих случаев?
5. Что такое "взвешенные ближайшие соседи" и как этот метод модифицирует классический k-NN? Какую проблему он помогает решить?
6. Объясните, почему нормализация признаков является критически важным этапом предобработки данных для метрических классификаторов. Что произойдет, если ее не провести?
7. В чем разница между минимаксной нормализацией и стандартизацией (Z-score normalization)? В каких ситуациях предпочтительнее использовать каждую из них?
8. Что такое кросс-валидация и для каких целей она применяется при использовании метода k ближайших соседей?
9. Опишите, как с помощью кросс-валидации можно подобрать оптимальное значение гиперпараметра k для алгоритма k-NN.
10. Каков главный computational drawback (вычислительный недостаток) наивной реализации k-NN? Какие существуют подходы и структуры данных для ускорения поиска ближайших соседей (назовите хотя бы два)?
11. В контексте лабораторной работы по классификации ирисов Фишера: как выбор метрики расстояния (Евклидова, Манхэттенская, Чебышева) мог повлиять на качество классификации? Обоснуйте, исходя из природы данных.
12. Представьте, что при классификации ирисов один из признаков (например, длина чашелистика) измерялся в миллиметрах, а другой (ширина лепестка) — в сантиметрах. Как это могло исказить результаты работы алгоритма k-NN и почему?
13. Что такое "проклятие размерности" и как оно проявляется в работе метрических алгоритмов, таких как k-NN, при большом количестве признаков?
14. Какие преимущества и недостатки метода k ближайших соседей по сравнению с другими алгоритмами классификации (например, с деревом решений) вы можете назвать?
15. По результатам лабораторной работы: какие выводы вы можете сделать о применимости метрических классификаторов для набора данных "Ирисы Фишера"? Насколько хорошо они справились с задачей?

УК-11.2. Применяет методики работы с этическими и социальными рисками, возникающими на разных стадиях жизненного цикла ИИ

Типовые тестовые вопросы

1. При работе с метрическим классификатором над данными о клиентах банка (включающими возраст, годовой доход в долларах и кредитный рейтинг) было решено применить Z-нормализацию. Какая ценностная предпосылка лежит в основе этого решения?

- a) Предпосылка, что форма распределения данных не важна для метрических алгоритмов.
- b) Предпосылка, что все признаки должны вносить равный вклад в вычисление расстояния, независимо от их исходного масштаба и единиц измерения.
- c) Предпосылка, что алгоритмы метрической классификации нечувствительны к выбросам.
- d) Предпосылка, что категориальные признаки всегда должны быть исключены из модели.

Правильный ответ: b)

2. Разрабатывая модель для прогнозирования оттока клиентов, аналитик использовал метод k-NN. Он один раз разделил данные на обучающую и тестовую выборки, добился на тестовой выборке точности в 95% и заключил: "Модель k-NN является оптимальным и надежным решением для этой задачи". Какое когнитивное искажение демонстрирует этот вывод?

- a) Искажение в результате поспешного обобщения (Hasty Generalization).
- b) Эффект знакомства (Familiarity Bias).
- c) Игнорирование базовой частоты (Base Rate Neglect).
- d) Иллюзия кластеризации (Clustering Illusion).

Правильный ответ: a)

3. Команда разрабатывает систему для автоматической сортировки резюме на вакансию "менеджер по продажам" в международной компании, используя метрический классификатор на основе данных об успешных сотрудниках из штаб-квартиры в стране X. Какая культурно-обусловленная предвзятость может быть заложена в такую постановку задачи?

- a) Алгоритм k-NN может быть вычислительно сложным для больших данных.
- b) Модель, обученная на данных из культуры X, может некорректно оценивать кандидатов из культуры Y, где стили коммуникации, образования и карьерные пути могут принципиально отличаться.
- c) Система может требовать сложной предобработки текстовых данных из резюме.
- d) Метод ближайших соседей плохо работает с категориальными признаками.

Правильный ответ: b)

4. Команда выбирает между стандартным k-NN и его взвешенной вариацией (Weighted k-NN). Они останавливаются на стандартном k-NN, аргументируя это так: "Все соседи должны иметь равный голос, это более демократично и справедливо". Какая ценностная предпосылка здесь ключевая?

- a) Предпосылка о линейной разделимости классов.
- b) Предпосылка о том, что простота реализации важнее точности.
- c) Предпосылка о равной важности всех наблюдений в окрестности, независимо от их фактической близости к целевому объекту.
- d) Предпосылка о том, что евклидова метрика является единственной верной.

Правильный ответ: c)

5. Сравнивая k-NN и метрический классификатор на основе одного и того же набора данных, исследователь получил следующие результаты: k-NN показал точность 89%, а другой метод — 85%. Исследователь заявляет: "k-NN объективно лучше для этого типа данных". Какое искажение может повлиять на этот вывод, если не были приняты определенные меры?

- a) Искажение выжившего (Survivorship Bias).
- b) Отклонение в сторону позитивного результата (Confirmation Bias).
- c) Отсутствие проверки статистической значимости различий (например, с помощью попарного t-теста на результатах кросс-валидации).
- d) Эффект ореола (Halo Effect).

Правильный ответ: c)

6. При развертывании системы реального времени для классификации изображений было решено использовать наивный полный перебор для поиска ближайших соседей, а не оптимизированные структуры данных (например, KD-деревья или LSH). Аргументация: "Так мы получим абсолютно точный результат и не исказим его приближенными вычислениями". Какая ценностная предпосылка здесь доминирует?

- a) Предпосылка о том, что вычислительные ресурсы неограничены и дешевы.
- b) Предпосылка о том, что точность вычисления расстояния всегда важнее скорости отклика системы.
- c) Предпосылка о том, что метрика Хэмминга лучше подходит для изображений, чем евклидова.
- d) Предпосылка о том, что не требуется предобработка данных.

Правильный ответ: b)

7. В задаче рекомендации фильмов используется k-NN на признаках "жанр", "режиссер" и "бюджет". Алгоритм находит пользователям "ближайших соседей" по их истории просмотров. Какая ключевая ценностная предпосылка встроена в саму постановку этой задачи?

- a) Предпосылка, что вкусы пользователей статичны и не меняются со временем.
- b) Предпосылка, что "похожесть" пользователей, определяемая выбранной метрикой расстояния и признаками, является адекватным заменителем совпадения их вкусов.
- c) Предпосылка, что все жанры имеют одинаковый вес при классификации.
- d) Предпосылка, что метод k-NN не требует настройки гиперпараметров.

Правильный ответ: b)

8. Международная команда собирает данные для медицинского диагност-ИИ, включающие рост и вес пациентов из разных стран. Перед использованием k-NN данные решено нормализовать методом Min-Max Scaling в диапазон [0,1]. Какая проблема может возникнуть?

- a) Min-Max Scaling не справляется с нормально распределенными данными.
- b) Глобальные минимумы и максимумы по всему набору данных могут быть заданы экстремальными значениями из специфических популяций, из-за чего нормализованные значения для основной массы пациентов окажутся в узком диапазоне, искажая понятие "расстояния".

- c) Этот метод нормализации неприменим к медицинским данным.
d) Метод k-NN становится неинтерпретируемым после такого преобразования.

Правильный ответ: b)

Типовые вопросы открытого типа:

1. Какие существуют методы борьбы с проблемой "проклятия размерностей" в метрических алгоритмах?
2. Как функция расстояния может быть адаптирована для работы с категориальными признаками, а не только с числовыми?
3. В чем заключаются ключевые различия между евклидовой, манхэттенской и метрикой Чебышева? В каких практических ситуациях каждая из них может быть предпочтительнее?
4. Опишите принцип работы алгоритма радиус-окрестности (Radius Neighbors). В чем его преимущества и недостатки по сравнению с k-NN?
5. Как можно оценить качество работы метрического классификатора, используя только обучающую выборку, без отложенной тестовой выборки?
6. Какие эвристические методы существуют для автоматического подбора оптимального значения k в алгоритме k-NN?
7. Как влияет наличие шума и выбросов в данных на работу метода k ближайших соседей?
8. Что такое "мера сходства" и чем она отличается от "метрики расстояния"? Приведите примеры мер сходства.
9. Какие методы предобработки данных особенно важны для метрических классификаторов при работе с разнородными признаками?
10. Как можно использовать взвешивание признаков для улучшения качества классификации в метрических алгоритмах?
11. В чем заключаются основные вычислительные сложности при использовании метрических классификаторов на больших объемах данных?
12. Какие существуют подходы к сокращению объема хранимых данных (редукции обучающей выборки) для метода k-NN без значительной потери точности?
13. Как можно адаптировать метрические методы классификации для решения задач регрессии?
14. Какие особенности набора данных "Ирисы Фишера" делают его удобным для демонстрации работы метрических классификаторов?
15. Какие метрики качества классификации наиболее уместно использовать для оценки работы метрических алгоритмов на задаче классификации ирисов и почему?

Код компетенции	Результаты освоения ОПОП Содержание компетенций
ОПК-6	Способен использовать в педагогической деятельности научные основы знаний в сфере информационно-коммуникационных технологий

ОПК-6.1. Демонстрирует понимание современного состояния информационно-коммуникационных технологий

Типовые тестовые вопросы

1. Какова основная цель кластеризации в машинном обучении?

- a) Прогнозирование точного значения целевой переменной.
- b) Разделение данных на группы (кластеры) таким образом, чтобы объекты внутри одной группы были похожи, а объекты из разных групп — отличались.
- c) Снижение размерности данных для визуализации.
- d) Определение наиболее значимых признаков для прогнозирования.

Правильный ответ: b)

2. Какой из перечисленных функционалов качества кластеризации предназначен для оценки среднего попарного сходства внутри кластеров и между ними?

- a) Индекс Данна (Dunn Index)
- b) Сумма квадратов внутрикластерных отклонений (WCSS)
- c) Силуэт (Silhouette Score)
- d) Критерий информативности (Information Criterion)

Правильный ответ: c)

3. Какой ключевой параметр необходимо задать перед запуском алгоритма K-средних (K-Means)?

- a) Порог останова для итеративного процесса.
- b) Начальные центроиды для всех кластеров.
- c) Количество кластеров K.
- d) Функцию расстояния между объектами.

Правильный ответ: c)

4. Какое утверждение лучше всего описывает разницу между иерархическими и статистическими алгоритмами кластеризации?

- a) Иерархические алгоритмы всегда точнее статистических.
- b) Статистические алгоритмы строят древовидную структуру (дендrogramму), а иерархические — нет.
- c) Иерархические алгоритмы создают вложенные разбиения, а статистические обычно предполагают, что данные порождены смесью распределений.
- d) Статистические алгоритмы не требуют задания числа кластеров.

Правильный ответ: c)

5. Что представляет собой дендрограмма в иерархической кластеризации?

- a) График зависимости качества кластеризации от числа кластеров.
- b) Диаграмму рассеяния данных в пространстве первых двух главных компонент.
- c) Древовидную структуру, показывающую последовательное объединение или разделение кластеров на всех уровнях.
- d) Визуализацию распределения данных внутри одного кластера.

Правильный ответ: c)

6. Алгоритм EM (Expectation-Maximization), используемый для Gaussian Mixture Models, является примером:

- a) Иерархического агломеративного алгоритма.

- b) Графового алгоритма.
- c) Статистического алгоритма, основанного на максимизации правдоподобия.
- d) Чисто эвристического метода без теоретического обоснования.

Правильный ответ: c)

7. Какой из перечисленных алгоритмов классифицируется как эвристический алгоритм кластеризации?

- a) EM-алгоритм для смеси гауссовых распределений.
- b) DBSCAN, основанный на плотности.
- c) Агломеративная иерархическая кластеризация.
- d) Кластеризация на основе моделей (Model-Based Clustering).

Правильный ответ: b)

8. Какая из перечисленных мер является примером функционала качества, основанного на внутренней оценке (без использования эталонных меток)?

- a) Adjusted Rand Index (ARI)
- b) Fowlkes-Mallows Index (FMI)
- c) Сумма квадратов внутрикластерных отклонений (WCSS)
- d) Однородность (Homogeneity)

Правильный ответ: c)

9. Какой принцип лежит в основе алгоритма DBSCAN?

- a) Минимизация расстояния от точек до центроидов их кластеров.
- b) Объединение самых близких кластеров на каждом шаге.
- c) Идея о том, что кластеры — это области высокой плотности данных, разделенные областями низкой плотности.
- d) Предположение, что данные порождены смесью нескольких распределений.

Правильный ответ: c)

10. Какой подход к выбору оптимального числа кластеров K для алгоритма K-средних использует метод "локтя" (Elbow Method)?

- a) Анализ дендрограммы.
- b) Поиск K, при котором стабилизируется индекс силуэта.
- c) Построение графика зависимости суммы квадратов внутрикластерных отклонений (WCSS) от K и поиск точки излома.
- d) Выбор K, который максимизирует вероятность данных при заданной модели.

Правильный ответ: c)

Типовые вопросы открытого типа:

1. В чем заключается принципиальное различие между задачами классификации и кластеризации?
2. Опишите пошаговый алгоритм работы K-Means. Каковы его ключевые преимущества и недостатки?
3. Что такое функционал качества кластеризации и для каких целей он используется? Назовите несколько примеров.

4. Как метод "локтя" (Elbow Method) помогает определить оптимальное количество кластеров? На что именно мы смотрим на графике?
5. В чем разница между агломеративным и дивизивным подходами в иерархической кластеризации?
6. Что такое дендрограмма и какую информацию мы можем из нее извлечь?
7. Объясните разницу между методами "одиночной" (Single), "полной" (Complete) и "средней" (Average) связи. Как выбор метода влияет на форму получаемых кластеров?
8. Каковы основные идеи, лежащие в основе алгоритма кластеризации по плотности DBSCAN?
9. Какие объекты в DBSCAN считаются объектами ядра, пограничными объектами и шумовыми выбросами?
10. В чем заключаются основные преимущества DBSCAN по сравнению с K-Means?
11. Что такое EM-алгоритм и как он применяется в кластеризации на основе смеси распределений?
12. Какое ключевое предположение о данных делает Гауссова смесь распределений (GMM) в отличие от K-Means?
13. Что такое "внутренние" и "внешние" метрики качества кластеризации? Приведите примеры и объясните, в каких ситуациях они применяются.
14. Опишите проблему, известную как "эффект цепочки" (chaining effect). Для какого метода иерархической кластеризации он наиболее характерен?
15. В чем состоит принцип "нечеткой кластеризации" (Fuzzy Clustering), например, в алгоритме C-Means? Чем он отличается от "жестких" методов, таких как K-Means?

ОПК-6.2. Применяет в педагогической деятельности информационно-коммуникационные технологии

Типовые тестовые вопросы

1. Какой из перечисленных алгоритмов кластеризации НЕ требует заранее задавать количество кластеров?

- a) K-средних (K-Means)
- b) EM-алгоритм для смеси гауссовых распределений
- c) Иерархическая кластеризация
- d) Алгоритм Fuzzy C-Means

Правильный ответ: c)

2. Что характеризует "эффект цепочки" (chaining effect) в иерархической кластеризации?

- a) Слишком быстрое образование очень больших кластеров
- b) Объединение кластеров по принципу ближайшего соседа может приводить к вытянутым, нежелательным кластерам
- c) Потеря отдельных объектов в процессе кластеризации
- d) Неправильное определение начальных центроидов

Правильный ответ: b)

3. Какой функционал качества кластеризации вычисляет среднее значение для каждого объекта отношения разности расстояний до ближайшего чужого кластера и среднего расстояния внутри своего кластера?

- a) Индекс Данна
- b) Calinski-Harabasz Index

- c) Силуэт (Silhouette Score)
- d) Davies-Bouldin Index

Правильный ответ: c)

4. Какой тип меры близости между кластерами используется в иерархической кластеризации при методе полной связи (complete linkage)?

- a) Расстояние между ближайшими объектами разных кластеров
- b) Расстояние между самыми дальными объектами разных кластеров
- c) Среднее расстояние между всеми объектами разных кластеров
- d) Расстояние между центроидами кластеров

Правильный ответ: b)

5. Какой из алгоритмов кластеризации наиболее устойчив к выбросам и может обнаруживать кластеры произвольной формы?

- a) K-средних
- b) Иерархическая кластеризация
- c) DBSCAN
- d) EM-алгоритм для смеси гауссовых распределений

Правильный ответ: c)

6. Что представляет собой ковариационная матрица в гауссовой модели смеси (GMM)?

- a) Коэффициенты важности каждого признака
- b) Форму и ориентацию кластера в пространстве признаков
- c) Среднее значение кластера
- d) Вероятность принадлежности точки к кластеру

Правильный ответ: b)

7. Какой из перечисленных методов является дивизивным (разделяющим) алгоритмом кластеризации?

- a) AGNES (Agglomerative Nesting)
- b) DIANA (Divisive Analysis)
- c) DBSCAN
- d) OPTICS

Правильный ответ: b)

8. Какой параметр в алгоритме DBSCAN определяет минимальное количество точек в ϵ -окрестности для того, чтобы точка считалась ядром?

- a) eps
- b) min_samples
- c) metric
- d) leaf_size

Правильный ответ: b)

9. Какой метод иерархической кластеризации использует расстояние между средними значениями кластеров?

- a) Одиночная связь (single linkage)

- b) Полная связь (complete linkage)
- c) Средняя связь (average linkage)
- d) Центроидный метод (centroid linkage)

Правильный ответ: d)

10. Какой из перечисленных алгоритмов позволяет точке принадлежать нескольким кластерам одновременно?

- a) K-средних
- b) DBSCAN
- c) Иерархическая кластеризация
- d) Нечеткая кластеризация (Fuzzy C-Means)

Правильный ответ: d)

Типовые вопросы открытого типа:

1. Какие основные проблемы возникают при использовании K-Means для кластеров несферической формы?
2. Как выбор метрики расстояния влияет на результат иерархической кластеризации?
3. В чем заключаются основные вычислительные сложности агломеративной иерархической кластеризации при работе с большими данными?
4. Какие параметры необходимо задать для алгоритма DBSCAN и как они влияют на результат кластеризации?
5. Как можно интерпретировать ковариационные матрицы в Гауссовой смеси распределений (GMM)?
6. В чем разница между минимизацией внутрикластерной дисперсии (K-Means) и максимизацией правдоподобия (GMM)?
7. Какие существуют методы инициализации центроидов в K-Means и как они влияют на конечный результат?
8. Как работает алгоритм BIRCH и в чем его преимущества для кластеризации больших datasets?
9. Что такое "силуэтный анализ" и как он помогает оценить качество кластеризации?
10. Какие существуют подходы к кластеризации категориальных данных, где неприменимы евклидовы расстояния?
11. Как можно использовать кластеризацию для предобработки данных в задачах классификации?
12. В чем заключаются основные ограничения иерархической кластеризации при работе с данными высокой размерности?
13. Как работает алгоритм OPTICS и чем он отличается от DBSCAN?
14. Какие существуют методы валидации результатов кластеризации при отсутствии эталонных меток?
15. Как можно адаптировать алгоритмы кластеризации для работы с потоковыми данными?

Код компетенции	Результаты освоения ОПОП Содержание компетенций
-----------------	---

ПК-19	Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ
-------	---

ПК-19.1. Позиционирует собственную задачу в заданной области знания с точки зрения трендов современного искусственного интеллекта

Типовые тестовые вопросы

1. Что из перечисленного является главным ограничением однослойного перцептрона Розенблатта?

- a) Неспособность обучаться методом градиентного спуска.
- b) Невозможность решить задачу, не являющуюся линейно разделимой.
- c) Отсутствие нелинейной функции активации.
- d) Требование большого объема вычислительных ресурсов даже для простых задач.

Правильный ответ: b)

2. Какова основная цель использования функции потерь (loss function) в процессе обучения модели?

- a) Увеличить скорость вычислений на графическом процессоре.
- b) Служить дифференцируемой мерой ошибки, которую необходимо минимизировать.
- c) Нормализовать входные данные перед подачей на вход нейрона.
- d) Определить архитектуру нейронной сети.

Правильный ответ: b)

3. В чем заключается ключевое (и зачастую неверное) предположение Наивного байесовского классификатора?

- a) Что данные распределены по нормальному закону.
- b) Что все наблюдения в выборке являются независимыми.
- c) Что все признаки являются условно независимыми при данном классе.
- d) что априорная вероятность класса $P(C)$ всегда равна 0.5.

Правильный ответ: c)

4. Какой компонент искусственного нейрона отвечает за введение нелинейности в его работу?

- a) Сумматор взвешенных входов.
- b) Смещение (bias).
- c) Функция активации.
- d) Весовые коэффициенты.

Правильный ответ: c)

5. Какой метод является фундаментальным для эффективного вычисления градиентов функции потерь по весам в глубокой нейронной сети?

- a) Метод наименьших квадратов.
- b) Метод обратного распространения ошибки (backpropagation).
- c) Прямой проход (forward pass).
- d) Наивный байесовский вывод.

Правильный ответ: b)

6. Если в задаче бинарной классификации два класса являются линейно разделимыми, что произойдет при обучении линейного SVM или перцептрона (при идеальных условиях)?

- a) Модель не сможет найти разделяющую гиперплоскость.
- b) Модель найдет гиперплоскость, которая обеспечит точность классификации 100% на обучающей выборке.
- c) Модель всегда будет переобучена.
- d) Модель будет минимизировать энтропию.

Правильный ответ: b)

7. Какая из следующих функций активации наиболее подвержена проблеме "исчезающего градиента"?

- a) ReLU (Rectified Linear Unit).
- b) Сигмоида.
- c) Софтмакс (Softmax).
- d) Линейная функция.

Правильный ответ: b)

8. Согласно теореме Байеса, как вычисляется апостериорная вероятность $P(\text{Class} | \text{Data})$?

- a) $P(\text{Data} | \text{Class}) * P(\text{Class})$
- b) $P(\text{Data} | \text{Class}) / P(\text{Data})$
- c) $[P(\text{Data} | \text{Class}) * P(\text{Class})] / P(\text{Data})$
- d) $P(\text{Data}) / P(\text{Class})$

Правильный ответ: c)

9. Какой из перечисленных алгоритмов оптимизации НЕ является эвристикой для улучшения сходимости градиентного спуска?

- a) Стохастический градиентный спуск (SGD).
- b) Момент.
- c) Adam.
- d) Метод опорных векторов (SVM).

Правильный ответ: d)

10. При сравнении линейных классификаторов на синтетических данных вы обнаружили, что логистическая регрессия работает значительно хуже, чем линейный SVM. Наиболее вероятная причина этого в том, что...

- a) Логистическая регрессия не поддерживает L2-регуляризацию.
- b) Классы являются линейно разделимыми, и SVM максимизирует зазор, что дает лучшую обобщающую способность.
- c) SVM по своей природе всегда точнее логистической регрессии.
- d) Логистическая регрессия не может работать с бинарными признаками.

Правильный ответ: b)

Типовые вопросы открытого типа:

ПК-19.2. Определяет тенденции развития, оценивает новизну и практическую значимость своих решений с точки зрения современного искусственного интеллекта

Типовые тестовые вопросы

1. Что такое принцип линейной разделимости и как он связан с построением разделяющей гиперплоскости?
2. Какие существуют типы функций потерь для задач классификации и регрессии и в чем их ключевые различия?
3. В чем заключается суть минимизации эмпирического риска и как это связано с проблемой переобучения?
4. Сформулируйте теорему Байеса и объясните ее значение для вероятностных моделей классификации.
5. В чем заключаются ключевые предположения наивного байесовского классификатора и как они влияют на его практическое применение?
6. Каковы основные компоненты искусственного нейрона и как они соотносятся с биологическим прототипом?
7. В чем заключались основные ограничения перцептрона Розенблатта и как они были преодолены в современных нейронных сетях?
8. Какие функции активации вы знаете и в чем их преимущества/недостатки для разных архитектур сетей?
9. Как метод обратного распространения ошибки вычисляет градиенты для многослойных нейронных сетей?
10. Какие проблемы могут возникать при использовании обычного градиентного спуска и как их решают современные оптимизаторы?
11. В чем разница между детерминистической и вероятностной постановками задачи классификации?
12. Какие эвристики используются для улучшения сходимости обучения нейронных сетей?
13. Как задача восстановления плотности распределения связана с байесовскими методами классификации?
14. Какие вычислительные возможности обеспечивают глубокие нейронные сети по сравнению с линейными моделями?
15. Как архитектура нейронной сети влияет на ее способность аппроксимировать сложные нелинейные зависимости?

1. Каков основной способ решения проблемы мультиколлинеарности в множественной линейной регрессии?

- a) Увеличение размера обучающей выборки
- b) Удаление одного из сильно коррелирующих признаков или использование регуляризации
- c) Применение нелинейных преобразований к признакам
- d) Нормализация целевой переменной

Правильный ответ: b)

2. Какой компонент сверточной нейронной сети (CNN) отвечает за уменьшение размерности карт признаков и повышение инвариантности к малым сдвигам?

- a) Сверточный слой (Convolutional layer)

- b) Слой подвыборки (Pooling layer)
- c) Полносвязный слой (Fully Connected layer)
- d) Функция активации ReLU

Правильный ответ: b)

3. Какая основная проблема решается с помощью механизмов LSTM и GRU в рекуррентных нейронных сетях?

- a) Высокая вычислительная сложность прямого распространения
- b) Проблема исчезающего и взрывающегося градиента при обучении на длинных последовательностях
- c) Неспособность работать с многомерными входными данными
- d) Отсутствие нелинейных преобразований

Правильный ответ: b)

4. Что минимизирует метод наименьших квадратов (МНК) в линейной регрессии?

- a) Сумму модулей разностей между предсказанными и истинными значениями
- b) Сумму квадратов разностей между предсказанными и истинными значениями
- c) Произведение стандартных отклонений признаков
- d) Корреляцию между остатками модели

Правильный ответ: b)

5. Какой тип нейронной сети наиболее подходит для обработки последовательностей переменной длины, таких как временные ряды или текст?

- a) Сверточная нейронная сеть (CNN)
- b) Полносвязная нейронная сеть (FC)
- c) Рекуррентная нейронная сеть (RNN)
- d) Автокодировщик (Autoencoder)

Правильный ответ: c)

6. Что является главным преимуществом использования сингулярного разложения (SVD) для решения задачи линейной регрессии?

- a) Возможность работы с категориальными признаками без их преобразования
- b) Численная устойчивость даже в случае вырожденных или плохо обусловленных матриц
- c) Автоматический подбор оптимальной степени полинома
- d) Невозможность переобучения модели

Правильный ответ: b)

7. Какой тип архитектуры нейронной сети наиболее эффективен для обработки изображений?

- a) Рекуррентная нейронная сеть (RNN)
- b) Полносвязная нейронная сеть (FC)
- c) Сверточная нейронная сеть (CNN)
- d) Ограниченная машина Больцмана (RBM)

Правильный ответ: c)

8. Какое утверждение лучше всего описывает проблему переобучения в линейной регрессии?

- a) Модель слишком проста и не может выявить закономерности в данных
- b) Модель слишком сложна и начинает учитывать шум в обучающих данных
- c) Модель не может работать с пропущенными значениями
- d) Модель требует нормализации всех входных признаков

Правильный ответ: b)

9. Какой из механизмов LSTM отвечает за "забывание" нерелевантной информации?

- a) Входная gate (input gate)
- b) Выходная gate (output gate)
- c) Gate забывания (forget gate)
- d) Слой активации tanh

Правильный ответ: c)

10. Какое утверждение о полносвязных слоях (FC) в нейронных сетях является верным?

- a) Они сохраняют пространственную структуру данных
- b) Каждый нейрон предыдущего слоя соединен с каждым нейроном следующего слоя
- c) Они эффективны для обработки последовательностей произвольной длины
- d) Они не содержат обучаемых параметров (весов)

Правильный ответ: b)

Типовые вопросы открытого типа:

1. Как линейная регрессия применяется в качестве базового элемента нейронных сетей?
2. В чем разница между простой и множественной линейной регрессией?
3. Как метод наименьших квадратов (МНК) связан с задачей оптимизации в линейной регрессии?
4. Каковы преимущества использования сингулярного разложения (SVD) для решения задачи линейной регрессии?
5. Как проблема мультиколлинеарности влияет на устойчивость оценок в линейной регрессии?
6. Какие методы regularization помогают бороться с переобучением в линейных моделях?
7. В чем особенности линейной монотонной регрессии и где она применяется?
8. Каковы основные преимущества и ограничения полносвязных нейронных сетей?
9. Как сверточные слои в CNN помогают выявлять пространственные закономерности в данных?
10. В чем заключаются основные проблемы обычных RNN при работе с длинными последовательностями?
11. Как механизм "вентиляй" в LSTM решает проблему исчезающего градиента?
12. Каковы ключевые различия между архитектурами LSTM и GRU?
13. Как методы pooling в CNN способствуют инвариантности к малым преобразованиям входных данных?
14. Какие преимущества дают современные фреймворки глубокого обучения по сравнению с ручной реализацией нейронных сетей?
15. Как выбор фреймворка глубокого обучения влияет на процесс разработки и deployment моделей?

Типовые теоретические вопросы для экзамена по дисциплине

1. Дайте строгое математическое определение метрики. Какие условия должна удовлетворять функция расстояния?
2. Опишите алгоритм k ближайших соседей и его основные модификации.
3. В чем заключается проблема "проклятия размерности" для метрических методов?
4. Какие методы ускорения поиска ближайших соседей вы знаете? Опишите принцип работы одного из них.
5. Объясните необходимость нормализации данных для метрических классификаторов.
6. Классификация алгоритмов кластеризации: иерархические, плотностные, центроидные методы.
7. Опишите алгоритм K-means и его ограничения.
8. В чем преимущества DBSCAN перед K-means? Когда его применение предпочтительнее?
9. Объясните принцип работы иерархической агломеративной кластеризации.
10. Какие метрики качества кластеризации вы знаете? Как выбрать подходящую метрику?
11. Принцип линейной разделимости и построение разделяющей гиперплоскости.
12. Объясните метод наименьших квадратов для линейной регрессии.
13. Проблема мультиколлинеарности и методы ее решения.
14. Теорема Байеса и ее применение в классификации.
15. Наивный байесовский классификатор: предположения и ограничения.
16. Сравнение биологического нейрона и его математической модели.
17. Опишите архитектуру и ограничения перцептрона Розенблatta.
18. Функции активации: сравнительный анализ и области применения.
19. Метод обратного распространения ошибки: основные шаги.
20. Проблема исчезающего градиента и методы ее решения.
21. Полносвязные сети: архитектура, преимущества и недостатки.
22. Сверточные нейронные сети: принцип работы и основные компоненты.
23. Объясните механизм работы рекуррентных нейронных сетей.
24. Архитектуры LSTM и GRU: сравнительный анализ.
25. Регуляризация в глубоких нейронных сетях: методы и их эффективность.
26. Принцип градиентного спуска и его модификации.
27. Сравнение SGD, Momentum, Adam оптимизаторов.
28. Функции потерь для задач классификации и регрессии.
29. Методы борьбы с переобучением в нейронных сетях.
30. Балансировка обучения: методы подбора learning rate.
31. Предобработка данных для нейронных сетей: основные техники.
32. Кросс-валидация и ее роль в оценке моделей.
33. Problem of vanishing/exploding gradients в глубоких сетях.
34. Transfer learning: принципы и практическое применение.
35. Инструменты и фреймворки глубокого обучения: сравнительный анализ.
36. Автокодировщики: архитектура и применение.
37. Generative Adversarial Networks: принцип работы.
38. Attention механизмы в нейронных сетях.
39. Transformer архитектура: основные компоненты.
40. Neural Architecture Search: современные подходы.
41. Анализ временных рядов с помощью RNN и LSTM.
42. Обработка естественного языка с использованием нейронных сетей.
43. Компьютерное зрение на основе CNN.
44. Рекомендательные системы на основе нейронных сетей.
45. Этические аспекты и bias в машинном обучении.