

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
ИМЕНИ В.Ф. УТКИНА»

Кафедра «Электронные вычислительные машины»

**Оценочные материалы  
Нейросетевые технологии и  
машинное обучение**

Направление подготовки  
09.03.01 «Информатика и вычислительная техника»

Направленность (профиль) подготовки  
«Программно-аппаратное обеспечение вычислительных  
комплексов и систем искусственного интеллекта»

Квалификация выпускника – бакалавр

Форма обучения – очная

## 1. Общеположения

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, защиты курсового проекта. Форма проведения экзамена - тестирование, письменный опрос по теоретическим вопросам и выполнение практических заданий.

## 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:**

| Шкалаоцени<br>вания                 | Критерий                                                                                                                   |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| 3 балла<br>(эталонныйуровень)       | уровень усвоения материала,<br>предусмотренного программой:<br>процент верных ответов на тестовые<br>вопросы от 85 до 100% |
| 2 балла<br>(продвинутыйуровень<br>) | уровень усвоения материала,<br>предусмотренного программой:<br>процент верных ответов на тестовые<br>вопросы от 70 до 84%  |
| 1 балл<br>(пороговыйуровень)        | уровень усвоения материала,<br>предусмотренного программой:                                                                |

|          |                                                                                                                    |
|----------|--------------------------------------------------------------------------------------------------------------------|
|          | процент верных ответов на тестовые вопросы от 50 до 69%                                                            |
| 0 баллов | уровень усвоения материала, предусмотренного программой:<br>процент верных ответов на тестовые вопросы от 0 до 49% |

**Описание критериев и шкалы оценивания теоретического вопроса:**

| <b>Шкала оценивания</b>          | <b>Критерий</b>                                                                                                                                                                |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 балла<br>(эталонный уровень)   | выставляется студенту, который дал полный ответ на вопрос, показал глубокие систематизированные знания, смог привести примеры, ответил на дополнительные вопросы преподавателя |
| 2 балла<br>(продвинутый уровень) | выставляется студенту, который дал полный ответ на вопрос, но на некоторые дополнительные вопросы преподавателя ответил только с помощью наводящих вопросов                    |
| 1 балл<br>(пороговый уровень)    | выставляется студенту, который дал неполный ответ на вопрос в билете и смог ответить на дополнительные вопросы только с помощью преподавателя                                  |
| 0 баллов                         | выставляется студенту, который не смог ответить на вопрос                                                                                                                      |

**Описание критериев и шкалы оценивания практического задания:**

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

**Описание критериев и шкалы оценивания курсового проекта**

| Шкала оц<br>ения                                     | Критерий                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Оценка<br>«отлично»<br>(эталонный уровень)           | курсовой проект выполнен в полном объеме, все аналитические этапы и модели выполнены без ошибок, дана оценка полученных результатов, работа выполнена самостоятельно, работа оформлена аккуратно, соблюдались сроки сдачи и защиты курсового проекта,<br>при защите курсового проекта студент ответил на все предложенные вопросы                                                                                           |
| Оценка<br>«хорошо»<br>(продвинутый уровень)          | курсовой проект выполнен в полном объеме, присутствуют незначительные ошибки при проведении анализа и/или при построении моделей, дана оценка полученных результатов, работа выполнена самостоятельно, работа оформлена аккуратно, соблюдались сроки сдачи и защиты курсового проекта, при защите курсового проекта студент ответил не на все предложенные вопросы (правильных ответов не менее 75%)                        |
| Оценка<br>«удовлетворительно»<br>(пороговый уровень) | курсовой проект выполнен в полном объеме, присутствуют ошибки при проведении анализа и/или при построении моделей, оценка полученных результатов не является полной, работа выполнена самостоятельно, по оформлению работы имеются замечания,<br>частично соблюдались сроки сдачи и защиты курсового проекта, при защите курсового проекта студент ответил не на все предложенные вопросы (правильных ответов не менее 50%) |
| Оценка<br>«неудовлетворительно»                      | курсовой проект выполнен не в полном объеме, присутствуют грубые ошибки при проведении анализа и/или при построении моделей, отсутствует оценка полученных результатов, работа выполнена не самостоятельно, по оформлению работы имеются замечания, не соблюдались сроки сдачи и защиты курсового                                                                                                                           |

|  |                                                                                                                     |
|--|---------------------------------------------------------------------------------------------------------------------|
|  | проекта, при защите курсового проекта студент ответил не на все предложенные вопросы (правильных ответов менее 50%) |
|--|---------------------------------------------------------------------------------------------------------------------|

*На промежуточную аттестацию в форме экзамена выносятся тест, два теоретических вопроса и 2 задачи. Максимально студент может набрать 15 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 15 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 10 до 14 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 5 до 9 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 5 баллов или не выполнил всех предусмотренных в течение семестра практических заданий.

#### **ЗПАСПОРТОЦЕНОЧНЫХМАТЕРИАЛОВПО ДИСЦИПЛИНЕ**

| <b>Контролируемые разделы (темы) дисциплины</b>           | <b>Код контролируемой компетенции (или её части)</b> | <b>Вид, метод, форма оценочного мероприятия</b> |
|-----------------------------------------------------------|------------------------------------------------------|-------------------------------------------------|
| Раздел 1. Основы работы с данными на Python               | ПК-8                                                 | Экзамен                                         |
| Раздел 2. Изучение открытых платформ для работы с данными | ПК-8                                                 | Экзамен                                         |
| Раздел 3. Понятие линейной регрессии                      | ПК-11, ПК-12, ПК-16                                  | Экзамен                                         |
| Раздел 4. Кластерный анализ и деревья принятия решений    | ПК-11, ПК-12                                         | Экзамен                                         |

|                                                                    |                           |         |
|--------------------------------------------------------------------|---------------------------|---------|
| Раздел 5. Генетические алгоритмы                                   | ПК-12, ПК-16              | Экзамен |
| Раздел 6. Основы искусственных нейронных сетей                     | ПК-16                     | Экзамен |
| Раздел 7. Обзор средств для автоматизации задач машинного обучения | ПК-16                     | Экзамен |
| Раздел 8. Принципы построения оптимизаторов нейронных сетей        | ПК-16                     | Экзамен |
| Раздел 9. Слои нейронных сетей                                     | ПК-11, ПК-16              | Экзамен |
| Раздел 10. Квантование и экспорт нейронных сетей                   | ПК-12                     | Экзамен |
| Раздел 11. Рекуррентные нейронные сети                             | ПК-16                     | Экзамен |
| Раздел 12. Анализ текста                                           | ПК-16                     | Экзамен |
| Раздел 13. Промежуточная аттестация                                | ПК-8, ПК-11, ПК-12, ПК-16 | Экзамен |

#### 4 ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ

##### Промежуточная аттестация в форме экзамена

##### Промежуточная аттестация в форме экзамена

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций                                                                                               |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| ПК-1            | Способен применять статистические методы для анализа данных, валидации моделей машинного обучения и проведения экспериментов в области ИИ (MF-4) |

**ПК-1.1. Применяет статистические методы анализа и машинного обучения для решения задач анализа данных и проведения экспериментов на данных**

**Типовые тестовые вопросы:**

**1. Что является основной целью статистического анализа данных?**

- а) Создание пользовательского интерфейса
- б) Выявление закономерностей и характеристик данных
- в) Увеличение объема данных
- г) Сжатие файлов

**Правильный ответ: б**

**2. Какое значение чаще всего используется для оценки центральной тенденции выборки?**

- а) Медиана
- б) Дисперсия

- в) Корреляция
- г) Энтропия

**Правильный ответ: а**

**3. Что показывает дисперсия выборки?**

- а) Среднее значение признака
- б) Количество объектов в выборке
- в) Степень разброса данных относительно среднего значения
- г) Число признаков

**Правильный ответ: в**

**4. Как называется статистическая зависимость между двумя признаками?**

- а) Нормализация
- б) Корреляция
- в) Кластеризация
- г) Квантование

**Правильный ответ: б**

**5. Какой коэффициент чаще всего используется для оценки линейной зависимости между признаками?**

- а) Коэффициент Пирсона
- б) Коэффициент Жаккара
- в) Коэффициент Соренсена
- г) Коэффициент Танимото

**Правильный ответ: а**

**6. Для чего выполняется разделение данных на обучающую и тестовую выборки?**

- а) Для увеличения объема данных
- б) Для оценки способности модели обобщать данные
- в) Для уменьшения числа признаков
- г) Для хранения результатов

**Правильный ответ: б**

**7. Какой метод относится к задачам машинного обучения без учителя?**

- а) Линейная регрессия
- б) Логистическая регрессия
- в) K-Means
- г) Дерево решений

**Правильный ответ: в**

**8. Что понимается под генеральной совокупностью?**

- а) Полный набор объектов исследования
- б) Обучающая выборка
- в) Тестовая выборка
- г) Подмножество данных

**Правильный ответ: а**

**9. Какой статистический показатель наиболее чувствителен к выбросам?**

- а) Медиана
- б) Мода
- в) Среднее арифметическое
- г) Квартиль

**Правильный ответ: в**

**10. Для чего применяется визуализация данных при проведении экспериментов?**

- а) Для выявления закономерностей и аномалий
- б) Для уменьшения размера датасета
- в) Для обучения модели
- г) Для кодирования признаков

**Правильный ответ: а**

**Типовые вопросы открытого типа**

1. Что понимается под статистическим анализом данных?
2. Какие основные характеристики используются для описания выборки?
3. Что такое среднее значение, медиана и мода?
4. Для чего используется дисперсия и стандартное отклонение?
5. Что показывает коэффициент корреляции?
6. В чем различие между обучающей и тестовой выборками?
7. Какие задачи относятся к машинному обучению с учителем и без учителя?
8. Какие методы визуализации используются при анализе данных?
9. Для чего проводится предварительный анализ данных перед обучением модели?
10. Какие этапы включает проведение эксперимента в области машинного обучения?

**Типовые практические задания**

1. Загрузить набор данных средствами Pandas и выполнить расчет основных статистических характеристик признаков.
2. Построить гистограммы распределения числовых признаков и выполнить их анализ.
3. Рассчитать среднее значение, медиану, моду, дисперсию и стандартное отклонение для выбранных признаков.
4. Построить корреляционную матрицу признаков и определить наиболее связанные параметры.
5. Выявить выбросы в наборе данных и предложить способы их обработки.
6. Выполнить разделение набора данных на обучающую и тестовую выборки.
7. Провести сравнительный анализ двух признаков с использованием коэффициента корреляции Пирсона.
8. Построить диаграммы рассеяния для анализа взаимосвязей между признаками.
9. Выполнить исследовательский анализ данных (EDA) и подготовить краткий аналитический отчет.
10. Провести полный статистический анализ набора данных и сформулировать выводы о его структуре, качестве и пригодности для построения моделей машинного обучения.

**ПК-1.2. Способен применять статистические методы для оценки качества моделей ИИ, включая метрики и критерии для регрессии, классификации и кластеризации, а также для проведения статистических тестов для сравнения моделей**

**Типовые тестовые вопросы:**

**1. Какая метрика показывает долю правильно классифицированных объектов?**

- а) Recall

- б) Accuracy
- в) F1-score
- г) Precision

**Правильный ответ: б**

**2. Какая метрика характеризует долю правильно определенных положительных объектов среди всех объектов, отнесенных моделью к положительному классу?**

- а) Precision
- б) Accuracy
- в) ROC-AUC
- г) RMSE

**Правильный ответ: а**

**3. Что показывает метрика Recall?**

- а) Долю правильно найденных объектов положительного класса
- б) Среднее значение ошибки
- в) Размер тестовой выборки
- г) Скорость обучения модели

**Правильный ответ: а**

**4. Для чего используется F1-score?**

- а) Для оценки времени обучения
- б) Для объединенной оценки Precision и Recall
- в) Для оценки качества кластеризации
- г) Для оценки объема данных

**Правильный ответ: б**

**5. Какая метрика чаще всего используется для оценки качества регрессии?**

- а) Accuracy
- б) Precision
- в) MSE
- г) ROC-AUC

**Правильный ответ: в**

**6. Что показывает коэффициент детерминации  $R^2$ ?**

- а) Степень объяснения моделью вариации целевой переменной
- б) Количество признаков модели
- в) Размер обучающей выборки
- г) Скорость работы алгоритма

**Правильный ответ: а**

**7. Для чего используется матрица ошибок (Confusion Matrix)?**

- а) Для хранения результатов экспериментов
- б) Для анализа качества классификации
- в) Для обучения модели
- г) Для балансировки выборки

**Правильный ответ: б**

**8. Какая метрика применяется для оценки качества кластеризации?**

- а) Silhouette Score
- б) MAE

в) Precision

г) Recall

**Правильный ответ: а**

**9. Что означает p-value в статистическом тестировании?**

а) Размер выборки

б) Вероятность получения наблюдаемого результата при верности нулевой гипотезы

в) Количество параметров модели

г) Скорость работы алгоритма

**Правильный ответ: б**

**10. Для чего применяется критерий Стьюдента?**

а) Для оценки качества изображений

б) Для сравнения средних значений выборок

в) Для кластеризации данных

г) Для уменьшения размерности

**Правильный ответ: б**

**Типовые вопросы открытого типа**

1. Какие основные метрики используются для оценки качества классификационных моделей?
2. В чем различие между Accuracy, Precision и Recall?
3. Для чего используется F1-score и когда его применение предпочтительно?
4. Какие метрики применяются для оценки регрессионных моделей?
5. Что показывает коэффициент детерминации  $R^2$ ?
6. Какую информацию содержит матрица ошибок?
7. Какие метрики используются для оценки качества кластеризации?
8. Что такое статистическая гипотеза и как выполняется ее проверка?
9. Для чего используются статистические критерии сравнения моделей машинного обучения?
10. Какие ограничения имеют различные метрики оценки качества моделей ИИ?

**Типовые практические задания**

1. Построить матрицу ошибок для результатов работы классификатора и выполнить ее анализ.
2. Рассчитать Accuracy, Precision, Recall и F1-score по результатам классификации.
3. Выполнить сравнительный анализ нескольких моделей классификации по основным метрикам качества.
4. Построить ROC-кривую и вычислить значение ROC-AUC.
5. Оценить качество регрессионной модели по метрикам MSE, RMSE, MAE и  $R^2$ .
6. Провести анализ влияния дисбаланса классов на значения Accuracy, Precision и Recall.
7. Выполнить кластеризацию данных и рассчитать показатель Silhouette Score.
8. Провести статистическое сравнение результатов двух моделей с использованием критерия Стьюдента.
9. Реализовать программный модуль вычисления метрик качества средствами Scikit-Learn.
10. Подготовить отчет по результатам сравнения нескольких моделей машинного обучения с использованием статистических методов оценки качества.

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций                                    |
|-----------------|---------------------------------------------------------------------------------------|
| ПК-8            | Способен применять язык программирования Python для решения задач в области ИИ (PL-1) |

**ПК-8.1** Разрабатывает и отлаживает прикладные решения разной сложности и для разного круга конечных пользователей с использованием языка программирования Python, тестирует, испытывает и оценивает качество таких решений

**ПК-8.2** Осуществляет выбор инструментов разработки на Python, приемлимых для создания прикладной системы обработки научных данных, машинного обучения и визуализации с заданными требованиями

**ПК-8.3** Разрабатывает и поддерживает системы обработки больших данных различной степени сложности

**Типовые тестовые вопросы:**

- Какой тип данных Pandas используется для работы с таблицами?
  - np.array
  - DataFrame**
  - List
  - Dictionary
- Какая метрика чаще всего используется для оценки точности модели в задаче бинарной классификации?
  - MeanAbsoluteError (MAE)
  - Accuracy (Доля правильных ответов)**
  - R2-Score
  - F1-Score
- Какая команда используется для установки библиотеки scikit-learn через pip?
  - pipscikit-learn
  - installscikit-learn
  - pythoninstallsklearn
  - pipinstallscikit-learn**
- Какой метод в Pandas используется для проверки наличия пропущенных значений в данных?
  - .describe()
  - .isnull().sum()**
  - .head()
  - .plot()
- Какой модуль Python используется для работы с многомерными массивами и математическими функциями?
  - pandas
  - math**

- в) numpy**  
г) os
6. Если ваша главная задача — быстро протестировать гипотезу с помощью классических алгоритмов ML (логистическая регрессия, случайный лес), какой фреймворк вы выберете?  
а) PyTorch  
б) TensorFlow  
**в) Scikit-learn**  
г) Keras
7. Какой фреймворк глубокого обучения известен своим динамическим вычислительным графом и популярен в исследовательской среде?  
а) Scikit-learn  
б) TensorFlow  
**в) PyTorch**  
г) Pandas
8. Если вам нужна библиотека для работы с тензорами и автоматического дифференцирования, но вы хотите более высокоуровневый API, чем "чистый" TensorFlow, что вы выберете?  
а) NumPy  
**б) Keras**  
в) Matplotlib  
г) OpenCV
9. Для какой цели чаще всего используется формат ONNX?  
а) Для визуализации данных  
**б) Для преобразования (экспорта) модели, обученной в одном фреймворке, для запуска в другом**  
в) Для очистки данных  
г) Для написания документации
10. Какой инструмент НЕ является библиотекой для машинного обучения?  
а) Scikit-learn  
**б) Pandas**  
в) XGBoost  
г) LightGBM
11. Какой структуры данных библиотеки NumPy следует избегать при работе с очень большими массивами из-за неэффективности?  
а) numpy.array  
**б) Стандартный список Python (list)**  
в) numpy.ndarray  
г) pandas.Series
12. Что такое «пропущенное значение» (missingvalue) в наборе данных?  
а) Значение, которое равно нулю.  
**б) Значение, которое не записано (обозначается как NaN или None).**

- в) Самое большое значение в столбце.
  - г) Выброс (аномалия).
13. Какой метод Pandas позволяет удалить строки или столбцы, содержащие пропущенные значения?
- а) `.fillna()`
  - б) `.dropna()`**
  - в) `.replace()`
  - г) `.isna()`
14. Что такое «выброс» (outlier) в данных?
- а) Первая строка таблицы.
  - б) Значение, которое значительно отличается от большинства других наблюдений.**
  - в) Название столбца.
  - г) Тип диаграммы.
15. Какой инструмент из `scikit-learn` используется для масштабирования числовых признаков (например, к диапазону `[0, 1]`)?
- а) `LabelEncoder`
  - б) `StandardScaler` или `MinMaxScaler`**
  - в) `train_test_split`
  - г) `RandomForestClassifier`

### Типовые теоретические вопросы

1. Вопрос: Напишите строчку кода на Python для импорта библиотеки `numpy` с псевдонимом `np`.

Ответ: `import numpy as np`

2. Вопрос: Что делает метод `.fit()` у модели в `scikit-learn` или `Keras`?

Ответ: Запускает процесс обучения модели

3. Вопрос: Как с помощью библиотеки `matplotlib` построить простой линейный график? Опишите ключевые шаги.

Ответ: `import matplotlib.pyplot as plt`

`ax.plot([какие-либо данные])`

`plt.show()`

Программный код может значительно отличаться в зависимости от используемого источника данных.

4. Вопрос: Для чего используется метод `train_test_split` из `sklearn.model_selection`?

Ответ: Метод позволяет случайным образом разделить выборку на обучающую и тестовую

5. Вопрос: Назовите основные библиотеки Python, которые вы будете использовать для полного цикла задачи ИИ (от данных до модели).

Ответ: изучаемый в программе стек – numpy, pandas, keras (допустимы ответы tensorflow, torch, matplotlib, seaborn, sklearn)

6. Вопрос: Назовите ключевое различие между Scikit-learn и PyTorch/TensorFlow. Для каких типов задач больше подходит каждый?

Ответ: первый предназначен для исследования гипотез и построения простейших моделей. Второй нацелен на разработку глубоких нейронных сетей.

7. Вопрос: В чём основное преимущество использования Keras вместе с TensorFlow?

Ответ: более компактный и понятный синтаксис, более быстрое вхождение в разработку

8. Если вы делаете проект по компьютерному зрению, какие основные библиотеки (кроме фреймворков DL) вам, скорее всего, понадобятся для загрузки и первичной обработки изображений?

Ответ: PIL и/или OpenCV. Допустимыми ответами являются (SKImage, FFMPEG)

9. Вопрос: Какой фактор, кроме типа задачи, может повлиять на выбор фреймворка?

Ответ: допустимы один или несколько вариантов, наиболее важных для студента, например: сообщество, производительность, легкость развертывания и другие.

10. Вопрос: Почему для обработки больших данных в Python используют структуры numpy.ndarray и pandas.DataFrame, а не стандартные списки?

Ответ: производительность. Описанные структуры используют низкоуровневые оптимизации для более быстрой обработки инструкций

11. Вопрос: Назовите два типичных способа обработки пропущенных значений в данных.

Ответ: Удаление строк с пропусками, замена пропусков значениями, близкими к реальным

12. Вопрос: Что такое "кодирование категориальных признаков" и зачем оно нужно? Приведите пример простого кодирования.

Ответ: Такое кодирование необходимо, поскольку большинство моделей работают с векторными представлениями признаков. Самый простой пример: One-Hot, когда номер класса меняется на единицу в целевом разряде вектора.

13. Как называется процесс преобразования "сырых" данных в формат, пригодный для обучения модели?

Ответ: Предобработка

14. Какие два метода в Pandas вы используете в первую очередь для первичного знакомства с данными (просмотр и общая статистика)?

Ответ: Head и Info

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций                                                                                                     |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| ПК-11           | Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками (ML-2) |

**ПК-11.1. Различает основные типы задач машинного обучения и применяет на практике принципы их решения**

**ПК-11.2. Применяет методы предварительной обработки данных и работы с признаками**

**Типовые тестовые вопросы:**

- Если модель предсказывает категорию объекта (например, "спам"/"не спам"), это задача:
  - Регрессии
  - Классификации**
  - Кластеризации
  - Ассоциации
- Какой тип задачи НЕ требует размеченных данных для обучения?
  - Классификация
  - Регрессия
  - Кластеризация**
  - Прогнозирование
- Целевая переменная (target) — это:
  - Самый важный признак
  - Данные, которые нужно очистить
  - То, что модель должна предсказать**
  - Гиперпараметр модели
- Метод MinMaxScaler используется для:
  - Кодирования текстовых признаков
  - Удаления пропусков
  - Масштабирования чисел в диапазон [0, 1]**
  - Поиска выбросов
- Метрика R2-Score (коэффициент детерминации) используется для оценки:
  - Классификационных моделей
  - Кластеризации
  - Регрессионных моделей**
  - Всех типов моделей
- Если в данных есть столбец "Цвет" со значениями "Красный", "Синий", "Зеленый", какой метод кодирования создаст отдельные бинарные столбцы?
  - LabelEncoding
  - OrdinalEncoding

- в) **One-HotEncoding**
  - г) Нормализация
7. Какая библиотека Python является основным инструментом для предобработки табличных данных?
- а) NumPy
  - б) Matplotlib
  - в) **Pandas**
  - г) Scikit-learn
8. Что такое train\_test\_split в Scikit-learn?
- а) Алгоритм кластеризации
  - б) Метод масштабирования данных
  - в) **Функция для разделения данных на обучающую и тестовую выборки**
  - г) Метрика качества
9. Если модель показывает точность 100% на обучающих данных и 60% на тестовых, это проблема:
- а) Недообучения (underfitting)
  - б) **Переобучения (overfitting)**
  - в) Нормального обучения
  - г) Случайности
10. Для задачи обнаружения мошеннических операций, где 99% операций легальны, важнее всего:
- а) Ассигасу (доля правильных ответов)
  - б) Precision (точность)
  - в) **Recall (полнота)**
  - г) Все метрики одинаково важны

### Типовые теоретические вопросы

1. Вопрос: Назовите три основных типа задач машинного обучения.

Ответ: классификация, регрессия, кластеризация

2. Вопрос: В чем разница между задачами классификации и регрессии?

Ответ: классификация определяет принадлежность к классу (категоризация), а регрессия работает с непрерывными данными.

3. Вопрос: Что такое "признак" (feature) в машинном обучении? Приведите пример для задачи оценки стоимости квартиры.

Ответ: Это отдельная характеристика или свойство данных, которые используются моделью для обучения и предсказания. То есть это элементы, которые описывают данные и на основе которых алгоритм делает выводы. Количество комнат, расстояние до центра и др.

4. Вопрос: Для чего нужна тестовая выборка?

Ответ: для финальной проверки корректности работы обученной модели

5. Вопрос: Назовите две основные проблемы, которые решает предобработка данных.

Ответ: удаление аномальных выбросов, приведение данных к диапазону.

6. Вопрос: Что такое выбросы (outliers) и почему они могут быть проблемой?

Ответ: Выбросы – это аномальные значения в данных. Чаще всего они вызваны ошибками в сборе датасета. Аномалии дают снижение точности работы нейронных сетей во время обучения.

7. Вопрос: Зачем нужно масштабировать числовые признаки перед обучением некоторых моделей (например, метода опорных векторов или градиентного спуска)?

Ответ: Устранение перекоса в важности признаков, ускорение обучения, увеличение стабильности работы алгоритмов

8. Вопрос: Объясните разницу между StandardScaler и MinMaxScaler.

Ответ: Первый позволяет выровнять данные, распределённые по нормальному закону распределения, второй полезен тогда, когда данные распределены равномерно.

9. Вопрос: Что такое пайплайн (pipeline) в Scikit-learn и зачем он нужен?

Ответ: это инструмент, который позволяет объединить несколько операций обработки данных в единую модель. Нужен для упрощения процесса обучения.

10. Вопрос: Назовите две метрики для оценки качества бинарной классификации, кроме Accuracy.

Ответ: Precision, Recall

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций                                                                             |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------|
| ПК-12           | Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения (ML-3) |

**ПК-12.1. Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи**

**ПК-12.2. Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ**

**Типовые тестовые вопросы:**

- Какой алгоритм из Scikit-learn используется для линейной регрессии?
  - LinearRegression**
  - LogisticRegression
  - KMeans
  - DecisionTreeRegressor
- Что делает алгоритм K-Means?
  - Предсказывает числовое значение
  - Классифицирует объекты по классам
  - Разбивает данные на кластеры**
  - Находит ассоциативные правила
- RandomForestClassifier — это:
  - Одно дерево решений
  - Ансамбль деревьев решений**
  - Линейная модель
  - Метод кластеризации
- Какая метрика НЕ используется для оценки регрессии?
  - Среднеквадратическая ошибка (MSE)
  - Accuracy**
  - R2-Score
  - Средняя абсолютная ошибка (MAE)
- Что такое перекрестная проверка (cross-validation)?
  - Метод увеличения данных
  - Способ более надежной оценки качества модели**
  - Алгоритм предобработки
  - Метод визуализации
- Какой алгоритм хорошо работает с категориальными признаками без специального масштабирования?
  - Линейная регрессия

- б) Метод опорных векторов (SVM)
  - в) Деревья решений**
  - г) К-ближайших соседей (KNN)
7. Метод fit() в Scikit-learn используется для:
- а) Предсказания на новых данных
  - б) Обучения модели**
  - в) Разделения данных
  - г) Оценки качества
8. Если зависимость между признаками нелинейная, какой алгоритм может ее уловить?
- а) Линейная регрессия
  - б) Полиномиальная регрессия**
  - в) Оба варианта
  - г) Ни один из вариантов
9. Что такое гиперпараметры модели?
- а) Веса, которые модель обучается на данных
  - б) Параметры, которые задаются до обучения (например, глубина дерева)**
  - в) Результаты предсказания
  - г) Признаки в данных
10. Для чего используется метод predict()?
- а) Для обучения модели
  - б) Для предсказания целевой переменной на новых данных**
  - в) Для оценки качества
  - г) Для визуализации данных

### Типовые теоретические вопросы

1. Вопрос: Назовите два простых алгоритма для задачи регрессии.

Ответ: линейная и логистическая регрессия

2. Вопрос: Что такое дерево решений и как оно принимает решение?

Ответ: это метод визуализации и анализа, используемый для систематизации и упрощения процесса принятия решений. Он представляет информацию в виде древовидной структуры с корнем, ветвями и узлами, где каждый узел представляет собой вопрос или условие, а ветви — возможные ответы или исходы.

3. Вопрос: В чем основная идея метода К-ближайших соседей (KNN)?

Ответ: алгоритм классификации и регрессии, основанный на гипотезе компактности, которая предполагает, что расположенные близко друг к другу объекты имеют одинаковую принадлежность к одному и тому же кластеру.

4. Вопрос: Почему линейная регрессия может плохо работать, если признаки не масштабированы?

Ответ: такая модель не учитывает справедливый вклад каждого признака в процесс обучения.

5. Вопрос: Что такое ансамбли моделей и в чем их преимущество? Назовите один ансамблевый метод.

Ответ: это метод, при котором несколько базовых моделей (классификаторов или регрессоров) объединяются для получения более точных и надёжных предсказаний, чем способна обеспечить любая отдельная модель

6. Вопрос: Зачем нужно разделение данных на обучение и тест? Что будет, если тестировать модель на тех же данных, на которых она обучалась?

Ответ: это требуется для оценки качества обучения на независимых данных, которых не было в обучении. Это позволит сделать честное тестирование модели. В противном случае мы можем столкнуться с *overfitting* на обучающих данных и не сумеем его вовремя определить.

7. Вопрос: Опишите стандартные этапы полного цикла машинного обучения (ML-проекта).

Примерный ответ: сбор данных, предварительная обработка и очистка данных, выбор модели, обучение модели, получение оценки качества работы модели, дальнейшее дообучение модели (при необходимости), экспорт модели, развертывание модели в целевой среде.

8. Вопрос: Как выбрать между простой линейной моделью и сложной моделью для конкретной задачи?

Ответ: если зависимости целевых переменных от входных данных тривиальны или есть на это подозрение – есть смысл выбирать малые модели. Если нет, то следуют увеличивать сложность модели с осторожностью. Слишком глубокие модели требуют больше ресурсов, склонны к переобучению и требуют датасеты больших объёмов.

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций                                                             |
|-----------------|----------------------------------------------------------------------------------------------------------------|
| ПК-16           | Способен проводить передовые исследования в области архитектур, алгоритмов МО, оптимизации и математики (FC-1) |

**ПК-16.1. Разрабатывает фундаментальные основы и новые алгоритмы машинного обучения**

**ПК-16.2. Разрабатывает новые архитектуры глубоких нейросетей**

**Типовые тестовые вопросы:**

- Какая функция активации чаще всего используется на выходном слое для задачи классификации?
  - ReLU
  - Softmax**
  - Tanh
  - Linear
- Какой алгоритм оптимизации является адаптивным и часто используется по умолчанию для обучения нейронных сетей?
  - SGD (Stochastic Gradient Descent)
  - Adam**
  - Градиентный спуск (Batch Gradient Descent)
  - RMSProp
- Что вычисляет функция потерь (loss function)?
  - Скорость обучения модели
  - Архитектуру нейронной сети
  - Меру ошибки между предсказанием модели и правильным ответом**
  - Количество параметров модели
- Какой процесс лежит в основе обучения глубоких нейронных сетей?
  - Прямое распространение (Forward propagation)
  - Обратное распространение ошибки (Backpropagation)**
  - Нормализация данных
  - Валидация модели
- Что такое эпоха (epoch) в процессе обучения?
  - Один шаг обновления весов
  - Полный проход по всему обучающему набору данных**
  - Проход по тестовому набору данных
  - Тип слоя в нейронной сети
- Какой тип слоя специально создан для обработки данных с пространственной структурой, например, изображений?
  - Полносвязный слой (Dense)
  - Свёрточный слой (Convolutional)**

- в) Рекуррентный слой (LSTM/GRU)
  - г) Слой Dropout
7. Какой слой обычно используется для уменьшения размерности карт признаков (featuremaps) в сверточной сети?
- а) Слой активации (Activation)
  - б) Слой пулинга (Pooling), например, MaxPooling**
  - в) Батч-нормализация (BatchNormalization)
  - г) Входной слой (Input)
8. Как называется архитектура нейронной сети, где выходы одного слоя подаются не только на следующий слой, но и на более поздние или сами на себя?
- а) Полносвязная сеть (Feed-Forward)
  - б) Сверточная сеть (CNN)
  - в) Рекуррентная сеть (RNN) или сеть с остаточными связями (ResNet)**
  - г) Автокодировщик (Autoencoder)
9. Какой простой слой помогает бороться с переобучением, случайным образом "отключая" часть нейронов во время обучения?
- а) Слой активации ReLU
  - б) Слой Dropout**
  - в) Слой BatchNormalization
  - г) Выходной слой (Output)
10. Если вы проектируете сеть для обработки текста или временных рядов, какой тип слоя будет основным?
- а) Свёрточный (Convolutional)
  - б) Полносвязный (Dense)
  - в) Рекуррентный (RNN/LSTM)**
  - г) Пулинг (Pooling)

### Типовые теоретические вопросы

1. Вопрос: Для чего нужна функция активации в нейроне? Назовите две самые популярные.

Ответ: Добавление нелинейного преобразования. Sigmoid, Tanh (возможны ответы: Relu, Elu, Softmax)

2. Вопрос: Что такое скорость обучения (learningrate) и почему она важна?

Ответ: она определяет шаг при сходимости к экстремуму. Большой шаг может привести к перескоку целевой точки, малый – к очень медленной сходимости в ограниченной окрестности

3. Вопрос: Как называется самый простой тип слоя, где каждый нейрон связан со всеми нейронами предыдущего слоя?

Ответ: полносвязный слой (Denselayer)

4. Объясните простыми словами, что такое переобучение (overfitting) модели.

Ответ: Ситуация, когда нейронная сеть демонстрирует хорошие результаты на обучающем наборе, но не работает на всех остальных

5. Если ваша модель на обучающих данных показывает очень низкую ошибку, а на тестовых — очень высокую, в чём вероятная проблема и как с ней можно бороться?

Ответ: самая вероятная причина – переобучение. Для её решения необходимо выбрать менее глубокую архитектуру, увеличить набор данных или снизить количество эпох.

6. Из чего состоит простейшая архитектура полносвязной нейронной сети?

Ответ: из полносвязных слоев и слоёв активации.

7. Что делает слой BatchNormalization и какую проблему он помогает решить?

Ответ: слой выполняет нормализацию таким образом, чтобы среднее значение выхода было равно нулю, а стандартное отклонение – единице. Позволяет избегать слишком сильных градиентов, чем улучшает обучение

9. Как в Keras/TensorFlow можно создать свою собственную последовательную (Sequential) архитектуру? Покажите на примере 2-3 слоев.

Пример ответа:

```
model = Sequential()
```

```
model.add(Dense(3, activation='relu'))
```

```
model.add(Dense(2, activation='sigmoid'))
```

10. Вопрос: Почему при создании новой архитектуры важно начинать с простых моделей и постепенно их усложнять?

Ответ: большие архитектуры могут быть избыточны для целевой задачи. Они потребуют больше ресурсов и могут переобучаться.

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций         |
|-----------------|------------------------------------------------------------|
| ПК-20           | Способен применять автоматическое машинное обучение (ML-7) |

**ПК-20.1. Обосновывает способы и варианты применения алгоритмов автоматического машинного обучения в задачах ИИ, включая их преобразование и адаптацию к специфике задачи**

**Типовые тестовые вопросы:**

**1. Что понимается под AutoML?**

- а) Автоматизация выбора, настройки и обучения моделей машинного обучения
- б) Автоматическое создание баз данных
- в) Автоматическое написание программного кода
- г) Автоматическая разметка изображений

**2. Правильный ответ: а**

Какова основная цель применения AutoML?

- а) Увеличение объема обучающих данных
- б) Автоматизация подбора модели и ее гиперпараметров
- в) Замена вычислительных ресурсов
- г) Снижение объема оперативной памяти

**Правильный ответ: б**

**3. Что понимается под гиперпараметрами модели?**

- а) Параметры, определяемые в процессе обучения
- б) Параметры, задаваемые до начала обучения модели
- в) Результаты прогнозирования модели
- г) Метрики качества модели

**Правильный ответ: б**

**4. Для чего применяется поиск гиперпараметров (Hyperparameter Optimization)?**

- а) Для повышения качества модели
- б) Для уменьшения размера датасета
- в) Для хранения модели
- г) Для визуализации данных

**Правильный ответ: а**

**5. Что представляет собой квантование модели?**

- а) Увеличение числа параметров модели
- б) Преобразование параметров модели в представление с меньшей разрядностью
- в) Изменение архитектуры сети
- г) Добавление новых слоев

**Правильный ответ: б**

**6. Какой результат обычно достигается при квантовании модели?**

- а) Уменьшение объема памяти и ускорение вычислений
- б) Увеличение количества признаков
- в) Повышение разрешения изображений
- г) Увеличение размера модели

**Правильный ответ: а**

**7. Что такое Grid Search?**

- а) Метод полного перебора комбинаций гиперпараметров
- б) Метод кластеризации данных
- в) Метод снижения размерности
- г) Метод аугментации данных

**Правильный ответ: а**

**8. Какой подход позволяет исследовать случайные комбинации гиперпараметров?**

- а) PCA
- б) Random Search
- в) K-Means
- г) Bagging

**Правильный ответ: б**

**9. Для чего используется кросс-валидация в AutoML?**

- а) Для оценки качества модели на различных подвыборках данных

- б) Для хранения результатов обучения
- в) Для уменьшения числа признаков
- г) Для визуализации данных

**Правильный ответ: а**

#### **10. Что является критерием выбора модели в системах AutoML?**

- а) Только скорость обучения
- б) Только размер модели
- в) Значения целевых метрик качества и ресурсные ограничения
- г) Количество строк программного кода

**Правильный ответ: в**

#### **Типовые вопросы открытого типа**

1. Что представляет собой технология AutoML и какие задачи она решает?
2. Какие основные этапы включает процесс автоматического машинного обучения?
3. Для чего выполняется подбор гиперпараметров модели?
4. Какие методы оптимизации гиперпараметров используются в современных системах AutoML?
5. Что такое Grid Search и Random Search?
6. Как применяется кросс-валидация при автоматическом выборе моделей?
7. Что понимается под квантованием моделей машинного обучения?
8. Какие преимущества и ограничения имеет квантование моделей?
9. Какие факторы необходимо учитывать при выборе модели для практического применения?
10. В каких задачах использование AutoML наиболее эффективно?

#### **Типовые практические задания**

1. Выполнить подбор гиперпараметров модели классификации с использованием Grid Search.
2. Реализовать поиск оптимальных параметров модели методом Random Search.
3. Провести сравнение нескольких моделей машинного обучения с использованием средств AutoML.
4. Оценить влияние различных гиперпараметров на качество модели классификации.
5. Выполнить кросс-валидацию модели и проанализировать результаты.
6. Построить конвейер автоматического машинного обучения средствами Scikit-Learn.
7. Исследовать влияние квантования модели на точность и скорость работы.
8. Выполнить сравнительный анализ исходной и квантованной модели.
9. Подобрать оптимальную модель для заданного набора данных с использованием AutoML-инструментов.
10. Подготовить отчет по результатам автоматического подбора моделей и гиперпараметров с обоснованием выбора итогового решения.

**ПК-20.2. Применяет методы повышения устойчивости, надежности, безопасности алгоритмов автоматического ML для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ**

#### **Типовые тестовые вопросы:**

1. Что понимается под устойчивостью модели машинного обучения?

- а) Способность модели сохранять качество работы при изменении входных данных и условий эксплуатации
- б) Размер модели в памяти
- в) Скорость загрузки модели
- г) Количество параметров модели

**Правильный ответ: а**

**2. Для чего проводится мониторинг качества модели после внедрения?**

- а) Для увеличения объема обучающих данных
- б) Для контроля изменения качества работы модели во времени
- в) Для изменения архитектуры модели
- г) Для хранения результатов обучения

**Правильный ответ: б**

**3. Что понимается под дрейфом данных (Data Drift)?**

- а) Увеличение размера датасета
- б) Снижение скорости работы модели
- в) Изменение распределения входных данных по сравнению с обучающей выборкой
- г) Изменение формата хранения данных

**Правильный ответ: в**

**4. Для чего используется регуляризация?**

- а) Для увеличения числа признаков
- б) Для хранения модели
- в) Для аугментации данных
- г) Для уменьшения вероятности переобучения модели

**Правильный ответ: г**

**5. Что является одной из целей квантования модели?**

- а) Уменьшение вычислительных затрат при сохранении приемлемого качества
- б) Увеличение количества параметров модели
- в) Повышение разрешения изображений
- г) Увеличение размера модели

**Правильный ответ: а**

**6. Что понимается под надежностью модели машинного обучения?**

- а) Способность модели стабильно выдавать корректные результаты в течение длительного времени
- б) Размер обучающей выборки
- в) Скорость работы процессора
- г) Количество эпох обучения

**Правильный ответ: а**

**7. Какой подход используется для оценки качества модели на новых данных?**

- а) Увеличение количества признаков
- б) Тестирование на независимой выборке
- в) Удаление выбросов
- г) Изменение формата данных

**Правильный ответ: б**

**8. Для чего применяется ансамблирование моделей?**

- а) Для повышения устойчивости и качества прогнозирования

- б) Для уменьшения объема данных
- в) Для хранения результатов
- г) Для визуализации данных

**Правильный ответ: а**

**9. Что является важным элементом обеспечения безопасности моделей ИИ?**

- а) Контроль доступа к данным и моделям
- б) Увеличение числа признаков
- в) Уменьшение объема памяти
- г) Использование только одного алгоритма

**Правильный ответ: а**

**10. Какой показатель чаще всего используется для контроля качества модели после внедрения?**

- а) Значение целевой метрики качества на новых данных
- б) Размер программного кода
- в) Количество файлов проекта
- г) Частота обновления операционной системы

**Правильный ответ: а**

**Типовые вопросы открытого типа**

1. Что понимается под устойчивостью модели машинного обучения?
2. Какие факторы могут приводить к снижению качества модели после внедрения?
3. Что такое дрейф данных (Data Drift) и как он влияет на качество прогнозирования?
4. Какие методы используются для борьбы с переобучением моделей?
5. Для чего применяется регуляризация?
6. Какие способы повышения надежности моделей машинного обучения существуют?
7. Как организуется мониторинг качества модели в промышленной эксплуатации?
8. Какие преимущества дает использование ансамблей моделей?
9. Какие угрозы безопасности существуют для систем искусственного интеллекта?
10. Какие методы используются для обеспечения надежности и безопасности моделей AutoML?

**Типовые практические задания**

1. Провести анализ качества модели на обучающей и тестовой выборках и выявить признаки переобучения.
2. Исследовать влияние различных коэффициентов регуляризации на качество модели.
3. Выполнить мониторинг изменения качества модели при поступлении новых данных.
4. Провести анализ дрейфа данных между обучающей и текущей выборками.
5. Сравнить качество работы одиночной модели и ансамбля моделей.
6. Выполнить квантование модели и оценить изменение точности и производительности.
7. Разработать процедуру контроля качества модели после внедрения в эксплуатацию.
8. Исследовать устойчивость модели к шумам и изменениям входных данных.
9. Подготовить рекомендации по обеспечению надежности и безопасности модели машинного обучения.
10. Подготовить аналитический отчет по результатам исследования устойчивости, надежности и качества работы модели AutoML в условиях эксплуатации.

| Код компетенции | Результаты освоения ОПОП<br>Содержание компетенций                                   |
|-----------------|--------------------------------------------------------------------------------------|
| ПК-21           | Способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей (DL-1) |

**ПК-21.1.** Объясняет и применяет математические основы нейронных сетей, включая расчет градиентов, методы оптимизации и алгоритм обратного распространения ошибки для эффективного обучения моделей

**Типовые тестовые вопросы:**

**1. Что является основой обучения нейронной сети?**

- а) Расчет градиентов функции потерь
- б) Сжатие данных
- в) Сегментация изображений
- г) Кластеризация объектов

**Правильный ответ: а**

**2. Для чего используется функция потерь (Loss Function)?**

- а) Для хранения весов модели
- б) Для оценки ошибки модели
- в) Для визуализации данных
- г) Для генерации признаков

**Правильный ответ: б**

**3. Что показывает градиент функции потерь?**

- а) Размер обучающей выборки
- б) Скорость работы модели
- в) Направление наиболее быстрого изменения функции
- г) Количество нейронов в сети

**Правильный ответ: в**

**4. Какой алгоритм используется для вычисления градиентов в многослойной нейронной сети?**

- а) K-Means
- б) PCA
- в) Random Forest
- г) Обратное распространение ошибки (Backpropagation)

**Правильный ответ: г**

**5. В чем заключается идея метода градиентного спуска?**

- а) В поиске минимума функции потерь путем последовательного изменения параметров модели
- б) В уменьшении числа признаков
- в) В разделении данных на кластеры
- г) В генерации новых данных

**Правильный ответ: а**

**6. Какой параметр определяет величину шага при обновлении весов нейронной сети?**

- а) Momentum
- б) Batch Size

в) Learning Rate

г) Epoch

**Правильный ответ: в**

**7. Что может произойти при слишком большом значении Learning Rate?**

а) Модель может не сойтись к минимуму функции потерь

б) Модель будет обучаться бесконечно долго

в) Размер датасета увеличится

г) Число нейронов уменьшится

**Правильный ответ: а**

**8. Для чего используется функция активации в нейронной сети?**

а) Для хранения параметров модели

б) Для внесения нелинейности в модель

в) Для уменьшения размера датасета

г) Для вычисления точности модели

**Правильный ответ: б**

**9. Какой алгоритм оптимизации является развитием градиентного спуска и использует адаптивную настройку скорости обучения?**

а) KNN

б) Naive Bayes

в) Adam

г) PCA

**Правильный ответ: в**

**10. Что понимается под эпохой обучения (Epoch)?**

а) Одно обновление весов сети

б) Полный проход по всей обучающей выборке

в) Один нейрон сети

г) Один слой модели

**Правильный ответ: б**

**Типовые вопросы открытого типа**

1. Какие математические принципы лежат в основе работы искусственных нейронных сетей?
2. Что представляет собой функция потерь и какую роль она играет при обучении модели?
3. Что такое градиент и как он используется при оптимизации нейронной сети?
4. В чем заключается принцип работы алгоритма обратного распространения ошибки?
5. Как работает метод градиентного спуска?
6. Как влияет скорость обучения (Learning Rate) на процесс оптимизации модели?
7. Для чего используются функции активации и какие виды функций активации применяются на практике?
8. В чем особенности алгоритмов оптимизации SGD, RMSProp и Adam?
9. Какие причины могут приводить к проблемам сходимости при обучении нейронных сетей?
10. Какие способы используются для повышения эффективности обучения глубоких нейронных сетей?

**Типовые практические задания**

1. Рассчитать значение функции потерь для заданных прогнозов модели и фактических значений целевой переменной.
2. Вычислить градиент функции потерь для заданной функции и определить направление изменения параметров модели.
3. Выполнить один шаг градиентного спуска для заданного набора параметров нейронной сети.
4. Исследовать влияние различных значений Learning Rate на процесс оптимизации модели.
5. Реализовать алгоритм градиентного спуска для решения задачи линейной регрессии средствами Python.
6. Выполнить расчет обратного распространения ошибки для простой двухслойной нейронной сети.
7. Сравнить работу функций активации Sigmoid, Tanh и ReLU на тестовом наборе данных.
8. Исследовать динамику изменения функции потерь в процессе обучения модели.
9. Сравнить эффективность оптимизаторов SGD и Adam при обучении нейронной сети.
10. Подготовить аналитический отчет по результатам исследования методов оптимизации и алгоритма обратного распространения ошибки для выбранной модели.

**ПК-21.2. Реализует неглубокие нейронные сети (перцептроны, MLP), выбирает количество и размер слоёв, подходящие функции активации и функции потерь для решения задач классификации и регрессии**

**Типовые тестовые вопросы:**

**1. Что представляет собой перцептрон?**

- а) Алгоритм кластеризации
- б) Простейшую однослойную нейронную сеть
- в) Метод снижения размерности
- г) Ансамбль деревьев решений

**Правильный ответ: б**

**2. Для чего используется многослойный перцептрон (MLP)?**

- а) Для моделирования сложных нелинейных зависимостей
- б) Для хранения данных
- в) Для сжатия изображений
- г) Для сортировки файлов

**Правильный ответ: а**

**3. Какой слой нейронной сети непосредственно получает входные данные?**

- а) Выходной слой
- б) Скрытый слой
- в) Входной слой
- г) Слой нормализации

**Правильный ответ: в**

**4. Какой функцией активации чаще всего пользуются в скрытых слоях современных нейронных сетей?**

- а) Softmax
- б) Linear
- в) Sigmoid
- г) ReLU

**Правильный ответ: г**

**5. Какая функция активации обычно используется в выходном слое задачи бинарной классификации?**

- а) Sigmoid
- б) ReLU
- в) Tanh
- г) ELU

**Правильный ответ: а**

**6. Какая функция активации чаще всего применяется в выходном слое многоклассовой классификации?**

- а) ReLU
- б) Softmax
- в) Linear
- г) Leaky ReLU

**Правильный ответ: б**

**7. Какая функция потерь обычно используется для задачи регрессии?**

- а) Cross-Entropy
- б) Hinge Loss
- в) Mean Squared Error (MSE)
- г) Focal Loss

**Правильный ответ: в**

**8. Какая функция потерь наиболее часто применяется в задаче бинарной классификации?**

- а) Binary Cross-Entropy
- б) Mean Absolute Error
- в) Huber Loss
- г) MSE

**Правильный ответ: а**

**9. Что может привести к переобучению многослойного перцептрона?**

- а) Недостаточное количество эпох
- б) Слишком сложная модель относительно объема данных
- в) Малое количество нейронов
- г) Использование тестовой выборки

**Правильный ответ: б**

**10. Какой метод используется для уменьшения переобучения нейронной сети?**

- а) Увеличение числа слоев без ограничений
- б) Удаление обучающей выборки
- в) Dropout
- г) Удаление функции потерь

**Правильный ответ: в**

**Типовые вопросы открытого типа**

1. Что представляет собой перцептрон и какие задачи он способен решать?
2. Чем многослойный перцептрон отличается от однослойного?
3. Как определяется количество входных и выходных нейронов в сети?
4. Какие факторы влияют на выбор числа скрытых слоев и нейронов?

5. Для чего используются функции активации в нейронных сетях?
6. В чем различия между функциями активации Sigmoid, Tanh и ReLU?
7. Какие функции потерь применяются в задачах классификации?
8. Какие функции потерь используются в задачах регрессии?
9. Какие признаки свидетельствуют о переобучении нейронной сети?
10. Какие методы применяются для повышения качества обучения MLP?

### **Типовые практические задания**

1. Разработать однослойный перцептрон для решения задачи бинарной классификации.
2. Реализовать многослойный перцептрон (MLP) средствами TensorFlow или PyTorch.
3. Исследовать влияние количества скрытых нейронов на качество классификации.
4. Сравнить результаты работы функций активации Sigmoid, Tanh и ReLU на одном наборе данных.
5. Реализовать задачу регрессии с использованием многослойного перцептрона.
6. Подобрать оптимальную функцию потерь для задачи классификации и обосновать выбор.
7. Выполнить настройку архитектуры MLP путем изменения количества скрытых слоев.
8. Исследовать влияние метода Dropout на качество обобщения модели.
9. Провести сравнительный анализ нескольких конфигураций MLP по показателям качества.
10. Подготовить отчет по результатам проектирования, обучения и тестирования многослойного перцептрона с обоснованием выбора архитектуры, функций активации и функции потерь.

### **ПК-21.3. Применяет современные архитектуры глубоких сетей для решения различных задач, понимая их внутреннюю структуру и особенности обучения**

#### **Типовые тестовые вопросы:**

**1. Какая архитектура нейронных сетей наиболее часто используется для обработки изображений?**

- а) LSTM
- б) Transformer
- в) Сверточная нейронная сеть (CNN)
- г) Hopfield Network

**Правильный ответ: в**

**2. Для решения каких задач преимущественно применяются рекуррентные нейронные сети (RNN)?**

- а) Обработки последовательностей и временных рядов
- б) Кластеризации данных
- в) Сжатия изображений
- г) Сегментации объектов

**Правильный ответ: а**

**3. Какое преимущество имеет архитектура LSTM по сравнению с обычной RNN?**

- а) Уменьшение числа признаков

- б) Возможность учитывать долгосрочные зависимости в последовательностях
- в) Снижение размерности данных
- г) Повышение разрешения изображений

**Правильный ответ: б**

**4. Какая архитектура лежит в основе современных больших языковых моделей?**

- а) CNN
- б) Autoencoder
- в) GAN
- г) Transformer

**Правильный ответ: г**

**5. Для чего используются автоэнкодеры (Autoencoders)?**

- а) Для обнаружения объектов
- б) Для кодирования и восстановления данных
- в) Для кластеризации изображений
- г) Для управления агентами

**Правильный ответ: б**

**6. Какая архитектура состоит из генератора и дискриминатора?**

- а) ResNet
- б) Transformer
- в) GAN
- г) LSTM

**Правильный ответ: в**

**7. Что является основной задачей генеративно-сопоставительных сетей (GAN)?**

- а) Генерация новых данных, похожих на обучающую выборку
- б) Классификация изображений
- в) Снижение размерности данных
- г) Поиск выбросов

**Правильный ответ: а**

**8. Какой механизм является ключевым элементом архитектуры Transformer?**

- а) Pooling
- б) Self-Attention
- в) Batch Normalization
- г) Dropout

**Правильный ответ: б**

**9. Какая архитектура предназначена для обработки последовательностей с сохранением контекста?**

- а) Decision Tree
- б) K-Means
- в) CNN
- г) LSTM

**Правильный ответ: г**

**10. Что является одним из преимуществ архитектуры ResNet?**

- а) Использование остаточных связей для обучения глубоких сетей
- б) Отсутствие скрытых слоев
- в) Возможность работать только с текстом

г) Использование только линейных преобразований

**Правильный ответ: а**

#### **Типовые практические задания**

1. Какие основные классы архитектур глубоких нейронных сетей существуют?
2. В чем особенности сверточных нейронных сетей (CNN)?
3. Для каких задач используются рекуррентные нейронные сети (RNN)?
4. Какие преимущества предоставляет архитектура LSTM?
5. В чем заключается принцип работы Transformer?
6. Что такое механизм внимания (Attention)?
7. Как устроены генеративно-сопоставительные сети (GAN)?
8. Какие задачи решаются с помощью автоэнкодеров?
9. В чем преимущества архитектуры ResNet по сравнению с классическими CNN?
10. Какие критерии используются для выбора архитектуры нейронной сети под конкретную задачу?

#### **Типовые вопросы открытого типа**

1. Реализовать сверточную нейронную сеть для задачи классификации изображений.
2. Провести сравнительный анализ архитектур CNN и MLP на одном наборе данных.
3. Реализовать рекуррентную нейронную сеть для прогнозирования временного ряда.
4. Исследовать влияние архитектуры LSTM на качество прогнозирования последовательностей.
5. Разработать модель на основе Transformer для обработки текстовых данных.
6. Реализовать автоэнкодер для сжатия и восстановления данных.
7. Разработать генеративно-сопоставительную сеть и выполнить генерацию новых объектов.
8. Провести сравнительный анализ архитектур CNN, LSTM и Transformer для различных типов данных.
9. Исследовать влияние глубины сети на качество обучения модели.
10. Подготовить аналитический отчет по результатам выбора, настройки и применения современных архитектур глубоких нейронных сетей для решения прикладной задачи искусственного интеллекта.

**ПК-21.4. Разрабатывает и оптимизирует специализированные архитектуры для работы с изображениями, учитывая их уникальные свойства**

#### **Типовые тестовые вопросы:**

**1. Какая особенность отличает сверточные нейронные сети (CNN) от многослойных перцептронов?**

- а) Использование рекуррентных связей
- б) Использование операций свертки для выделения признаков
- в) Отсутствие скрытых слоев
- г) Использование только линейных функций

**Правильный ответ: б**

**2. Для чего применяется операция свертки в CNN?**

- а) Для извлечения локальных признаков изображения
- б) Для уменьшения объема памяти компьютера
- в) Для хранения изображений
- г) Для генерации новых данных

**Правильный ответ: а**

**3. Каково назначение слоя Pooling?**

- а) Увеличение числа признаков
- б) Выполнение классификации
- в) Уменьшение размерности карт признаков
- г) Хранение весов модели

**Правильный ответ: в**

**4. Какая архитектура считается одной из первых успешных сверточных нейронных сетей?**

- а) ResNet
- б) EfficientNet
- в) YOLO
- г) LeNet

**Правильный ответ: г**

**5. Какая архитектура впервые продемонстрировала существенный рост качества распознавания изображений на соревновании ImageNet?**

- а) U-Net
- б) AlexNet
- в) MobileNet
- г) Autoencoder

**Правильный ответ: б**

**6. В чем основное преимущество архитектуры ResNet?**

- а) Использование остаточных связей (Residual Connections)
- б) Отсутствие сверточных слоев
- в) Использование только одного слоя
- г) Полный отказ от функций активации

**Правильный ответ: а**

**7. Для какой задачи чаще всего используется архитектура U-Net?**

- а) Классификация изображений
- б) Сегментация изображений
- в) Кластеризация объектов
- г) Анализ временных рядов

**Правильный ответ: б**

**8. Какая архитектура ориентирована на детекцию объектов в режиме реального времени?**

- а) PCA
- б) LSTM
- в) YOLO
- г) K-Means

**Правильный ответ: в**

**9. Что является основным преимуществом MobileNet?**

- а) Высокая эффективность на мобильных и встраиваемых устройствах
- б) Использование рекуррентных слоев
- в) Работа только с текстовыми данными
- г) Полное отсутствие сверток

**Правильный ответ: а**

#### **10. Для чего применяется аугментация данных при обучении CNN?**

- а) Для увеличения разнообразия обучающей выборки и уменьшения переобучения
- б) Для уменьшения числа классов
- в) Для хранения изображений
- г) Для удаления шума из модели

**Правильный ответ: а**

#### **Типовые практические задания**

1. Какие принципы лежат в основе работы сверточных нейронных сетей?
2. Что такое операция свертки и какую роль она играет при обработке изображений?
3. Для чего используются слои Pooling?
4. В чем особенности архитектур LeNet и AlexNet?
5. Какие преимущества предоставляет архитектура VGG?
6. В чем заключается идея остаточных связей в ResNet?
7. Для каких задач применяется архитектура U-Net?
8. Какие особенности имеют модели семейства YOLO?
9. Какие методы используются для борьбы с переобучением CNN?
10. Какие критерии необходимо учитывать при выборе архитектуры для задачи анализа изображений?

#### **Типовые вопросы открытого типа**

1. Разработать сверточную нейронную сеть для классификации изображений на заданном наборе данных.
2. Реализовать архитектуру LeNet и провести ее обучение на тестовом наборе изображений.
3. Выполнить сравнительный анализ архитектур LeNet, AlexNet и ResNet по качеству классификации.
4. Исследовать влияние количества сверточных слоев на качество модели.
5. Реализовать аугментацию изображений и оценить ее влияние на результаты обучения.
6. Выполнить дообучение предобученной модели ResNet для прикладной задачи классификации изображений.
7. Реализовать сегментацию изображений с использованием архитектуры U-Net.
8. Настроить и применить модель YOLO для задачи детекции объектов.
9. Провести сравнительный анализ производительности моделей ResNet, MobileNet и EfficientNet на одном наборе данных.
10. Подготовить отчет по результатам проектирования, обучения, тестирования и оптимизации архитектуры компьютерного зрения с обоснованием выбора используемой модели.