

«УТВЕРЖДАЮ»
Декан факультета
компьютерной инженерии
_____ Т. Кучкоров
«___» _____ 2026 г.

**Министерство цифровых технологий Республики Узбекистан
Ташкентский университет информационных технологий имени
Мухаммада аль-Хорезми**

**ВОПРОСЫ ПО ПРЕДМЕТУ “MACHINE LEARNING”
ДЛЯ ПРОВЕДЕНИЯ ИК**

1. Дайте определение машинного обучения и объясните, чем оно отличается от классического программирования на примере.
2. Опишите полный цикл ML-проекта: от постановки задачи до внедрения модели, кратко охарактеризовав каждый этап.
3. Сравните контролируемое и неконтролируемое обучение на конкретных примерах задач.
4. Объясните, что такое задача классификации, регрессии и кластеризации, и приведите по одному примеру каждой.
5. Приведите пример задачи из вашей области и сформулируйте её в терминах машинного обучения (тип задачи, вход, выход).
6. Объясните различия между обучением с учителем, без учителя и с подкреплением по типу данных и обратной связи.
7. Приведите пример задачи, которая может быть решена как supervised, так и unsupervised способом, и сравните подходы.
8. Опишите роль награды (reward) в обучении с подкреплением и приведите пример среды агент-среда.
9. Объясните, когда имеет смысл комбинировать несколько видов обучения в одной системе, приведите пример.
10. Назовите типичные ошибки при неправильном выборе вида обучения для задачи и разберите один конкретный случай.
11. Объясните, чем отличаются обучающая, валидационная и тестовая выборки, и зачем нужно каждое разбиение.
12. Опишите шаги создания обучающего датасета «с нуля» для прикладной задачи (сбор, разметка, проверка качества).
13. Обсудите плюсы и минусы использования открытых датасетов, приведя примеры возможных рисков.
14. Перечислите основные шаги предобработки данных и поясните на одном примере, зачем каждый шаг нужен.
15. Приведите пример, когда некорректная предобработка данных приводит к неправильному выводу модели, и объясните, почему.
16. Опишите типичную структуру проекта по машинному обучению на Python: какие библиотеки и для чего используются.
17. Объясните преимущества и ограничения Google Colab для выполнения ML-экспериментов.

18. Сравните scikit-learn, TensorFlow и PyTorch по типичным задачам, для которых их применяют.
19. Кратко опишите pipeline «pandas + scikit-learn»: загрузка данных, предобработка, обучение, оценка, сохранение модели.
20. Объясните, когда и зачем использовать GPU при обучении моделей, и какие даёт это преимущества.
21. Объясните, как выборки и признаки представляются с помощью векторов и матриц в машинном обучении.
22. Покажите, как матричное умножение связано с вычислением выходов линейной модели для всей выборки.
23. Объясните геометрический смысл скалярного произведения и его использование в методах классификации.
24. Кратко опишите роль собственных значений и векторов в методах уменьшения размерности (например, PCA).
25. Как вы объясните важность линейной алгебры для понимания работы нейронных сетей студенту без сильной математики?
26. Дайте определение линейной регрессии и запишите её модель в общем виде, пояснив смысл коэффициентов.
27. Объясните, почему в линейной регрессии часто используется MSE как функция стоимости и какие у этого есть минусы.
28. Опишите алгоритм градиентного спуска для одномерной линейной регрессии: обновление параметров и критерий останова.
29. Сравните аналитическое решение (метод наименьших квадратов) и градиентный спуск по преимуществам и ограничениям.
30. Приведите ситуацию, где линейная регрессия систематически ошибается, и объясните, как это обнаружить по остаткам.
31. Объясните, чем множественная линейная регрессия отличается от простой, на примере задачи с несколькими признаками.
32. Дайте определение полиномиальной регрессии и покажите, как она позволяет моделировать нелинейные зависимости.
33. Объясните, как увеличение степени полинома влияет на переобучение и качество на тестовой выборке.
34. Покажите связь полиномиальных признаков с идеей «ядерных» методов (качественно, без формальных выводов).
35. Опишите подход к выбору признаков для множественной регрессии и проблему мультиколлинеарности.
36. Дайте определение задачи классификации и приведите примеры бинарной и многоклассовой классификации.
37. Сравните постановки задач классификации и регрессии по типу выходных данных и метрикам оценки.
38. Объясните проблему несбалансированных классов и приведите методы борьбы с ней.
39. Опишите основные шаги решения задачи классификации текста: от сырого текста до предсказания класса.
40. Приведите пример, когда высокая ассигасу вводит в заблуждение, и объясните, какие метрики использовать вместо неё.
41. Объясните принцип логистической регрессии и отличие её выхода от линейной регрессии.
42. Покажите роль логистической (sigmoid) функции в бинарной классификации и интерпретацию её выхода.

43. Объясните, почему для логистической регрессии используют $\log \text{loss}$, а не MSE, и что это даёт при обучении.
44. Опишите обучение логистической регрессии методом градиентного спуска и интерпретацию полученных весов.
45. Приведите пример практической задачи, в которой логистическая регрессия даёт конкурентное решение, и обоснуйте.
46. Объясните идею наивного байесовского классификатора и смысл «наивного» предположения.
47. Опишите алгоритм k-ближайших соседей и влияние параметра k на качество и устойчивость модели.
48. Кратко объясните идею SVM: что такое разделяющая гиперплоскость и максимальный зазор.
49. Опишите принцип построения решающего дерева: выбор признака, критерий разбиения, остановка роста дерева.
50. Объясните, как формируется случайный лес и почему ансамбль деревьев обычно работает лучше одного дерева.
51. Сравните k-NN, логистическую регрессию и SVM по интерпретируемости, скорости обучения и предсказания.
52. Объясните, какие проблемы возникают у «глубоких» решающих деревьев без ограничений и как их решает регуляризация.
53. Приведите пример задачи, где вы бы сравнили как минимум три классических алгоритма классификации, и опишите критерии выбора лучшего.
54. Объясните, как масштабирование признаков влияет на работу k-NN, SVM и логистической регрессии.
55. Опишите, как организовать честный эксперимент по сравнению нескольких классификаторов (разбиения, метрики, повторяемость).
56. Объясните понятия переобучения и недообучения и приведите по одному признаку каждого случая.
57. Сравните L1- и L2-регуляризацию по влиянию на веса модели и интерпретацию признаков.
58. Объясните, зачем нужны метрики качества и почему одной метрики часто недостаточно для оценки модели.
59. Приведите пример задачи, где выбор неправильной метрики приводит к неверной оценке модели, и предложите правильную метрику.
60. Перечислите и кратко охарактеризуйте практические методы борьбы с переобучением, помимо L1/L2-регуляризации.
61. Дайте определение обучения без учителя и объясните, зачем оно нужно при отсутствии меток.
62. Приведите примеры задач кластеризации, уменьшения размерности и обнаружения аномалий, кратко описав каждую.
63. Объясните назначение методов уменьшения размерности (например, PCA) и их роль в визуализации данных.
64. Дайте определение аномалии в данных и опишите подходы к её обнаружению методами ML.
65. Объясните, как результаты алгоритмов без учителя могут улучшить обучение с учителем (на примере кластеризации).
66. Подробно перечислите шаги алгоритма k-средних и опишите, по какому критерию он останавливается.
67. Объясните, как выбирается число кластеров k, и опишите метод «локтя» или силуэтный анализ.

68. Приведите пример данных, на которых k -средних даёт плохую кластеризацию, и объясните причину.
69. Сравните k -средних с иерархической кластеризацией по предположениям о форме кластеров и сложности.
70. Объясните, как оценивать качество кластеризации k -средних при отсутствии истинных меток.
71. Объясните различия агломеративной и дивизивной иерархической кластеризации.
72. Опишите процесс построения дендрограммы и способ выбора количества кластеров по ней.
73. Сравните разные меры расстояния (single-link, complete-link, average-link) в иерархической кластеризации и их влияние на результат.
74. Приведите пример, когда дендрограмма даёт более полезную информацию, чем просто «плоские» кластеры.
75. Объясните ограничения иерархической кластеризации по размеру данных и предложите способы их обхода.
76. Объясните интуитивно, что такое bias и variance, и как они связаны с пере- и недообучением.
77. Опишите процедуру k -fold кросс-валидации и её преимущества перед единственным разбиением train/test.
78. Объясните, как по матрице ошибок рассчитываются Precision, Recall и F1-мера и что они означают.
79. Сравните метрики MAE, MSE, RMSE и R^2 в регрессии и объясните, когда какую лучше использовать.
80. Опишите процесс выбора модели: сравнение нескольких алгоритмов, настройка гиперпараметров и принятие окончательного решения.
81. Опишите устройство искусственного нейрона: входы, веса, суммирование, активация, выход.
82. Объясните, какие задачи может решать однослойный перцептрон и какие принципиально не может.
83. Приведите примеры функций активации (sigmoid, ReLU, tanh) и сравните их свойства.
84. Объясните разницу между функцией потерь и функцией стоимости в контексте обучения нейросети.
85. Кратко опишите алгоритм градиентного спуска в нейросетях и роль параметров обучения (скорость, количество эпох).
86. Объясните, как нейронная сеть решает задачу регрессии и чем она отличается от нейросетевого классификатора.
87. Дайте качественное описание алгоритма обратного распространения ошибки (backpropagation).
88. Опишите простую архитектуру многослойной сети для классификации изображений малого размера и обоснуйте выбор.
89. Перечислите основные практические проблемы при обучении нейросетей и способы их смягчения.
90. Объясните, как корректно сравнивать нейросетевые модели с классическими алгоритмами на одной задаче.
91. Дайте определение глубокого обучения и объясните, чем «глубокая» сеть отличается от «мелкой».
92. Опишите этапы работы над задачей глубокого обучения: выбор архитектуры, подготовка данных, обучение, регуляризация, оценка.

93. Объясните роль скрытых слоёв в многослойной нейронной сети и почему их увеличение позволяет моделировать сложные зависимости.
94. Приведите по одному примеру использования глубокого обучения в компьютерном зрении, НЛП и рекомендациях, указав типичные архитектуры.
95. Объясните риски бессистемного увеличения глубины сети и приведите примеры негативных эффектов.
96. Опишите полный жизненный цикл реального ML-проекта в выбранной вами области: от постановки задачи до мониторинга модели.
97. Придумайте пример системы, где последовательно применяются unsupervised, supervised и reinforcement learning, и кратко опишите архитектуру.
98. Объясните, почему увеличение объёма и качества данных часто важнее усложнения модели, приведя пару примеров.
99. Опишите основные принципы этичного и ответственного применения машинного обучения (данные, конфиденциальность, fairness).
100. Сформулируйте пример учебного мини-проекта по машинному обучению (цель, данные, инструменты, результат), по которому можно оценить компетенции студента.

Заведующий кафедрой

Ответственный составитель



Б.Р. Азимов

А.Н. Шарифбаев