



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(ДГТУ)**

**ОЛИМПИАДА «Я – МАГИСТР»
ДЛЯ ПОСТУПАЮЩИХ В МАГИСТРАТУРУ в 2026 году**

09.04.02 Информационные системы и технологии/ Искусственный интеллект,
математическое моделирование и суперкомпьютерные технологии в разработке

информационных систем

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПОДГОТОВКИ
К ЗАКЛЮЧИТЕЛЬНОМУ ЭТАПУ ОЛИМПИАДЫ**

Составители:

Председатель методической комиссии:

доц., к.ф.-м.н. Мисюра Валентина Владимировна

Член методической комиссии:

доц., к.ф.-м.н. Богачева Марина Николаевна

ЗАКЛЮЧИТЕЛЬНЫЙ ЭТАП

Характер и уровень сложности олимпиадных задач направлены на достижение целей проведения Олимпиады: выявление и поддержка лиц, проявивших выдающиеся способности; стимулирование учебно-познавательной и научно-исследовательской деятельности обучающихся; развитие у обучающихся интеллектуальных и творческих способностей; создание необходимых условий для формирования качественного контингента магистрантов, ориентированных на продолжение академической карьеры; формирование системы непрерывного взаимодействия с одаренной и талантливой молодежью; распространение и популяризация научных знаний; привлечение талантливой молодежи, в том числе из зарубежных стран, к обучению в магистратуре.

Задания дифференцированы по сложности и требуют различных временных затрат на верное и полное решение. Задания направлены на выявление интеллектуального потенциала, аналитических способностей и креативности мышления участников и т.п.

Очный этап Олимпиады проводится только в письменной форме. Каждый участник Олимпиады получает бланк с заданием по варианту, содержащий 5 заданий. При выполнении заданий требуется:

1. Решения должны быть основаны на применении методов искусственного интеллекта, машинного обучения, методов теории вероятностей и математической статистики, теории алгоритмов, управления данными.

2. Алгоритмы и модели должны быть обоснованы с точки зрения их применимости к поставленной задаче.

3. Требуется оценка эффективности предлагаемого ИИ-решения (метрики качества, сравнение с традиционными подходами).

4. Материалы должны быть изложены корректным научно-техническим языком и сопровождаться обоснованными выводами.

При подготовке к Олимпиаде следует повторить приведенные ниже темы.

ПЕРЕЧЕНЬ ЭЛЕМЕНТОВ СОДЕРЖАНИЯ, ВКЛЮЧЕННЫХ В ЗАДАНИЯ ОЛИМПИАДЫ ЗАКЛЮЧИТЕЛЬНОГО ЭТАПА 2025/2026 УЧЕБНОГО ГОДА

Тема 1. Управление данными

Раздел посвящен методам проектирования модели данных с нормализацией, методам написания эффективных и правильных SQL-запросов (SELECT, INSERT, UPDATE, DELETE, JOIN и тд).

Пример 1.1

Для предметной области «Стрелковый клуб» приведите к третьей нормальной форме отношение «Тренировки по стрельбе», включающее следующие атрибуты:

Идентификатор (ПК),

ФИО клиента,

Телефон клиента,

Номер абонемента,
Модель оружия,
Серия и номер оружия,
ФИО инструктора,
Дата и время тренировки,
Стоимость аренды оружия.

Особенности предметной области:

- у каждого клиента один абонемент,
- абонемент принадлежит одному клиенту; каждое оружие принадлежит одному клиенту или стрелковому клубу,
- одному клиенту может принадлежать несколько единиц оружия;
- на каждой тренировке присутствует один клиент и один инструктор;
- на каждой тренировке клиент может использовать несколько единиц оружия, принадлежащих ему или клубу;
- стоимость аренды оружия, принадлежащего клубу, зависит от модели оружия. Результат представьте в виде схемы базы данных в одной из общепотребительных нотаций.

Для каждого отношения приведите перечень атрибутов с указанием первичных и внешних ключей.

Напишите на SQL следующие запросы:

1. Самые дорогие модели оружия.
2. Клиенты, которые арендовали оружие, принадлежащее клубу.

Ответ: представлен на рисунке 1.

Схема БД:



Запросы:

- 1) `select * from Модели_оружия where стоимость = (select max(стоимость) from Модели_оружия);`
- 2) `select * from Клиенты where номер_абонемента IN (select T.абонемент from Тренировки T, Используемое_оружие И, Оружие О where T.идентификатор = И.id_тренировки and И.id_оружия = О.серия_номер and О.владелец is null);`

Рисунок 1. Ответ на задачу по теме «Управление данными»

Тема 2. Алгоритмы и структуры данных

Раздел посвящен изучению методов организации данных и способов их эффективной обработки для решения вычислительных задач.

Основные направления этого раздела:

1. Структуры данных

Способы организации данных в памяти компьютера для обеспечения быстрого доступа и модификации:

- Линейные: массивы, списки (односвязные, двусвязные), стеки, очереди.
- Нелинейные: деревья (бинарные, деревья поиска, AVL), графы.
- Ассоциативные: хэш-таблицы, словари.

2. Алгоритмы

Последовательность действий для решения конкретных задач:

- Сортировка: пузырьком, быстрая (QuickSort), слиянием (MergeSort) и др.
- Поиск: линейный, бинарный.
- Алгоритмы на графах: поиск в ширину (BFS) и глубину (DFS), поиск кратчайшего пути (Дейкстра, A*).
- Рекурсия и динамическое программирование.

3. Оценка эффективности (Сложность)

Анализ как время выполнения и потребление памяти при исполнении алгоритма растут с увеличением объема данных.

- Big O Notation: нотация для оценки временной ($O(n)$, $O(\log n)$) и пространственной сложности.

Студенту рекомендуется при решении задачи обратить внимание на

- оптимизацию: написание алгоритма, который работает быстро и потребляет минимум ресурсов.
- масштабируемость: понимание того, как алгоритм поведет себя на миллионах строк данных.

Пример 2.1.

Дан массив из n целых чисел. Необходимо вычислить количество пар чисел в массиве, сумма которых равна нулю. Написать оптимальный по числу операций в терминах "O" (O большое) алгоритм подсчета и оценить его сложность. Размер памяти и диапазон целых чисел считать неограниченными.

Один из вариантов решения - алгоритма сложности $O(n)$ в среднем описан ниже:

1. Нам понадобится дополнительный контейнер - словарь (ассоциативный массив), реализованный на основе хэш-таблицы. Сложность операций поиска и добавления элементов в таком словаре $O(1)$ в среднем.

2. Проходим по массиву, выбирая лишь неотрицательные элементы и пишем в словарь, где ключ - значение элемента, значение по ключу - количество повторений. Это занимает линейную сложность.

2. Проходим по массиву, отбирая только отрицательные элементы. Проверяем наличие противоположных по знаку элементов в словаре. Если они есть, увеличиваем

счётчик пар с нулевой суммой на количество элементов в словаре. Это аналогично занимает линейную сложность.

3. Добавляем к счётчику количество пар из нулевых элементов (число нулевых элементов записано в словаре). Это занимает константную сложность ($O(1)$).

Пример 2.2.

В каталоге компьютера находится N файлов произвольного типа, каждый файл занимает не более, чем M байт. Необходимо определить количество уникальных по содержанию файлов, создать папку под каждую категорию и переместить туда одинаковые файлы. При этом в оперативную память компьютера возможно загрузить не более 5 файлов одновременно. Написать оптимальный по числу операций алгоритм и оценить его сложность в терминах "O" (O большое).

Алгоритм решения сложности $O(N*M)$ в среднем.

1. Будем интерпретировать каждый байт файла как символ, таким образом содержимое файла можно рассматривать как текст\строку.

2. Последовательно считываем каждый файл и формируем хеш таблицу, где ключ - содержимое файла, значение - название файла. Подсчет хеш-значения по содержимому файла требует сложности $O(M)$ для одного файла, для всех файлов - $O(N*M)$ соответственно.

3. При попытке добавить значение в хеш таблицу можно сразу определять наличие записи с данным ключом (поиск по хеш таблице требует средней сложности $O(1)$).

4. Если записи с таким ключом еще нет - создаем папку (именем папки может быть, например, значение хеша) перемещаем туда файл, если есть - перемещаем файл в уже существующую папку.

Для реализации алгоритма в зависимости от используемого языка программирования в качестве структуры данных могут быть использованы: `std::unordered_map`, `std::unordered_multimap` (C++), `dict` (Python), `HashMap` (Java).

Тема 3. Решение прикладных задач с помощью методов анализа и прогнозирования

Раздел посвящён методам извлечения полезной информации из накопленных данных для понимания текущей ситуации и предсказания будущего развития событий в конкретной сфере (бизнесе, технике, медицине и др.).

Ожидается, что студент владеет методами:

- статистического анализа (расчет средних значений, отклонений, корреляций и т.д.);
- интеллектуального анализа (поиск скрытых закономерностей, кластеризация);
- визуализации (построение графиков, дашбордов);
- анализа временных рядов (предсказание на основе истории, например, сезонность продаж).
- машинного обучения (ML) (использование нейросетей и алгоритмов (регрессия, деревья решений и т.д.) для прогнозов).

Пример 3.1

Сформулируйте задачу машинного обучения для прогнозирования превышения сроков ввода в эксплуатацию промышленного объекта.

Необходимо описать возможные источники данных, возможную структуру набора данных, целевую переменную, типы переменных, методы предобработки данных, предложить методы решения задачи, метрики качества.

Разбор задания

Ожидается, что студент:

- определит задачу как задачу классификации или регрессии;
- предложит использовать данные о сроках, типе объекта, подрядчике, погоде, BIM-параметрах и т.д. с указанием возможных типов переменных, укажет целевую переменную;
- опишет основные методы предобработки данных
- аргументирует выбор метрик (ROC-AUC, RMSE, MAE и др.).

Пример 3.2

Сформулируйте задачу машинного обучения для предсказания отказов промышленного оборудования.

Необходимо описать возможные источники данных, возможную структуру набора данных, целевую переменную, типы переменных, методы предобработки данных, предложить методы решения задачи, метрики качества.

Разбор задания

Ожидается, что студент:

- определит задачу как задачу классификации или логистической регрессии;
- предложит использовать данные о работе оборудования, например: температура, давление, расход, плановые/неплановые ремонты, заменённые детали, причины отказов, тип оборудования, возраст, конфигурация и т.д.;
- аргументирует выбор метрик (ROC-AUC, RMSE, MAE и др.);

Пример 3.3

Пусть имеется выборка x_i — количество решённых задач по машинному обучению, а y_i — количество чашек чая, выпитых при решении задач.

x_i	y_i
1	1
2	2
2	8

Рассмотрим модель $y_i = \beta x_i + u_i$. Функция потерь $\sum (y_i - \beta \hat{x}_i)^2$

- а) Найдите МНК-оценку β для имеющихся трёх наблюдений (Модель 1).
- б) Нарисуйте исходные точки и полученную прямую регрессии.

- в) Выведите формулу для оценки β в общем виде для n наблюдений.
 г) Оцените β с помощью функции потерь $\sum |y_i - \beta \hat{x}_i|$ (Модель 2)
 д) Оцените β с помощью функции потерь $\sum (y_i - \beta \hat{x}_i)^2 + 3\beta^2$ (Модель 3 с регуляризатором)

Решение:

- а) Выписываем ошибку и минимизируем её по параметрам модели:

$$(1 - \beta)^2 + (2 - 2\beta)^2 + (8 - 2\beta)^2 \rightarrow \min_{\beta}$$

Берём производную и приравниваем её к нулю:

$$\begin{aligned} -2(1 - \beta) - 4(2 - 2\beta) - 4(8 - 2\beta) &= 0, \\ (1 - \beta) + 2(2 - 2\beta) + 2(8 - 2\beta) &= 0, \\ 9\beta &= 21, \quad \text{оценка } \beta = 21/9. \end{aligned}$$

Итоговая модель для прогнозов: $y = (21/9)x$.

Обратите внимание, что в данном случае мы оценивали именно зависимость y от x . Для каждого x будет свой прогноз y .

- б) На рисунке 2 представлены наблюдения и оценённая прямая.

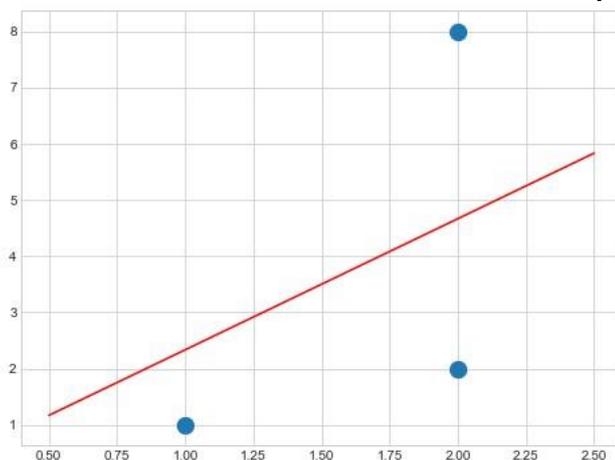


Рисунок 2. Исходные точки и полученная прямая регрессии

- в) Выведем формулу для случая n наблюдений. Ошибка будет иметь вид:

$$(y_1 - \beta x_1)^2 + (y_2 - \beta x_2)^2 + \dots + (y_n - \beta x_n)^2 \rightarrow \min_{\beta}$$

Берём производную, приравниваем её нулю и выписываем необходимую формулу оценки параметра модели.

г) Выписываем ошибку $|1 - \beta| + |2 - 2\beta| + |8 - 2\beta| \rightarrow \min_{\beta}$.

Изобразим на графике функцию потерь (рис. 3). Она будет кусочно-линейной.

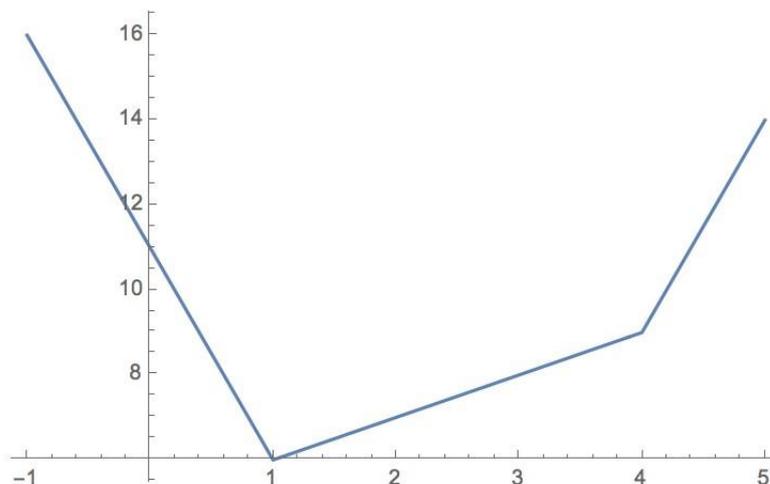


Рисунок 3. График функции потерь для исходных данных $\sum |y_i - \beta \hat{x}_i|$

Функция достигает минимума в точке 1, получается что оценка $\beta = 1$.

д) Модель с регуляризатором. Очевидно, что λ выбрана произвольно, без выполнения условия оптимальности.

$$(1 - \beta)^2 + (2 - 2\beta)^2 + (8 - 2\beta)^2 + 3\beta^2 \rightarrow \min_{\beta}$$

Берём производную и приравниваем ее нулю:

$$-2(1 - \beta) - 4(2 - 2\beta) - 4(8 - 2\beta) + 6\beta = 0,$$

$$(1 - \beta) + 2(2 - 2\beta) + 2(8 - 2\beta) - 3\beta = 0,$$

$$12\beta = 21, \quad \text{оценка } \beta = 21/12.$$

Видим, что из-за наличия регуляризатора, коэффициент β уменьшился.

Нарисуем все три линии регрессии в одной системе координат и немного порассуждаем о них.

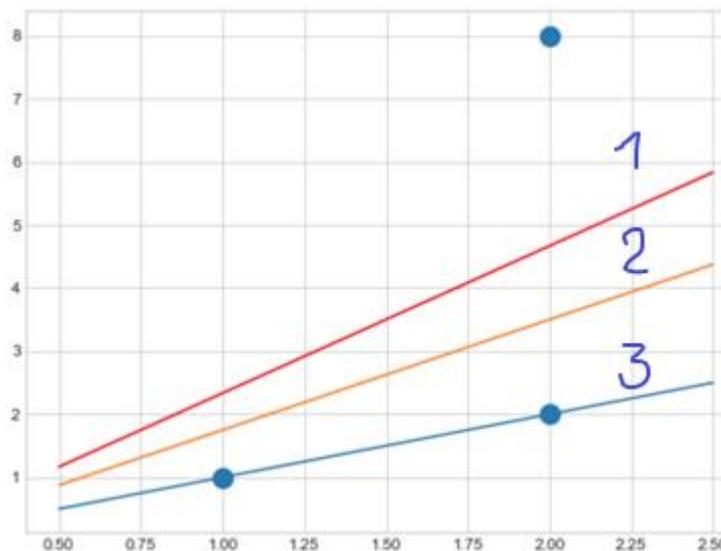


Рисунок 4. Графики моделей

Точку (2, 8) можно считать выбросом. Модель 1 довольно сильно под этот выброс подстраивается. Модель переобучается под выборку. Модель 3 более устойчива. Из-за регуляризатора она оказывается ниже Модели 1.

Модель 2 никак не реагирует на наличие в данных выброса. Это происходит из-за того, что минимизирует MAE, а не MSE как в остальных случаях. MSE в качестве прогноза выдаёт математическое ожидание. Оно чувствительно к выбросам. MAE в качестве прогноза выдаёт медиану. Она нечувствительна к выбросам. Если вы строите модель и подозреваете, что её очень сильно портят выбросы, попробуйте использовать в качестве функции ошибки MAE вместо MSE для её обучения.

Список источников для подготовки\

1. Гмурман, В. Е. Теория вероятностей и математическая статистика : учебник для вузов / В. Е. Гмурман. — 12-е изд. — Москва : Юрайт, 2023. — 479 с.
2. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы. Построение и анализ. 4-е изд. М.: Вильямс, 2023
3. Маркин, А. В. Программирование на SQL в 2 ч. Часть 2. Учебник и практикум для вузов. 3-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 385 с.
4. Мхитарян, В. С. Теория вероятностей и математическая статистика : учебник для вузов / В. С. Мхитарян, Е. В. Астафьева, Ю. Н. Миронкина, Л. И. Трошин ; под редакцией В. С. Мхитаряна. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 464 с.
5. Толстобров, А. П. Управление данными : учебное пособие для вузов / А. П. Толстобров. — Москва : Издательство Юрайт, 2024. — 272 с.
6. Соколов Е. А. Машинное обучение : материалы курса [Электронный ресурс]. — URL: <https://github.com/esokolov/ml-course-hse>
7. Учебник по машинному обучению // Яндекс Образование : [сайт]. — URL: <https://education.yandex.ru/handbook/ml>

**Методические рекомендации
по формированию критериев проверки (оценивания) олимпиадных
заданий заключительного этапа**

Заключительный этап олимпиады «Я – магистр» для поступающих в
магистратуру в 2026 году

Олимпиада по 09.04.02 Информационные системы и технологии/
Искусственный интеллект, математическое моделирование и суперкомпьютерные
технологии в разработке информационных систем
Критерии проверки.

1. Вариант преамбулы к критериям проверки заданий:

Вариант заключительного этапа Олимпиады по 09.04.02 Информационные системы и технологии/ Искусственный интеллект, математическое моделирование и суперкомпьютерные технологии в разработке информационных систем включает в себя 5 заданий разного типа. Каждое задание оценивается от 0 до 20 баллов. Наибольшая итоговая сумма баллов, которой могут быть оценены ответы на все вопросы олимпиадного варианта при условии отсутствия в них ошибок, неправильных, неполных или неточных ответов, равна 100. Неверные ответы оцениваются в 0 баллов. Возможен частичный зачёт баллов за неполный ответ на задание. Под неполным понимается ответ, содержащий правильные ответы не на все вопросы задания. В таком случае присуждается только часть баллов за правильные ответы задания, соответствующая доле от максимально возможного балла. Подсчёт итоговой оценки за задание осуществляется путём суммирования баллов, выставленных за каждый из вопросов.

1.1 Вариант 1 оформления критериев к олимпиадным заданиям:

Тип задания	Количество заданий в варианте	Критерий оценивания	Максимальное количество баллов за задание
Раздел (тема) 1. Управление данными			
Вопрос 1. Проектирование БД, написание запросов	1	Правильная схема БД (максимальная оценка 12баллов). За каждую ошибку снижение оценки на 1 балл Правильные SQL-запросы (максимальная оценка 8 баллов) Снижение на 2 балла за каждую ошибку запроса	20

Раздел (тема) 2. Алгоритмы и структуры данных			
Вопрос 1. Разработка алгоритма	1	<p>20 баллов: представлен корректный алгоритм сложности не более $O(N)$ в среднем с корректной оценкой памяти не более $O(N)$.</p> <p>16-18 баллов: представлен корректный алгоритм сложности не выше, чем $O(N)$ в среднем, но при описании алгоритма имеются некритические ошибки, не проанализирована используемая память.</p> <p>10-15 баллов: представлен корректный алгоритм сложности не выше, чем $O(N * \log N)$ в среднем, алгоритм описан не полностью, может быть неоднозначно истрактован, отсутствует оценка использования памяти.</p> <p>1-9 балла: представлен некорректный алгоритм, некоторые этапы или идеи которого верны.</p>	20
Раздел (тема) 3. Применение ИИ для анализа и прогнозирования в строительстве			
Вопрос 1. Постановка задачи машинного обучения	2	<p>20 баллов: правильно определен тип задачи с учетом типа переменных, входные переменные и целевая переменная, методы предобработки данных, определены метрики качества</p> <p>15 баллов: правильно определен тип задачи с учетом типа переменных, входные переменные и целевая переменная, методы предобработки данных, не определены метрики качества</p> <p>15 баллов: правильно определен тип задачи без учета типа переменных, входные переменные и целевая переменная, методы предобработки данных, определены метрики качества</p> <p>10 баллов: правильно определен тип задачи и входные переменные</p> <p>5 баллов: правильно определен тип задачи</p>	40
Вопрос 2. Методы машинного обучения	1	<p>20 баллов: правильно построено регрессионное дерево, выполнен прогноз, правильно обоснована возможность применения другой функции потерь</p> <p>15 баллов: правильно построено регрессионное дерево, выполнен прогноз, нет обоснования</p>	20

		<p>возможность применения другой функции потерь</p> <p>15 баллов: не доведено до конца построение регрессионного дерева, не выполнен прогноз, правильно обоснована возможность применения другой функции потерь</p> <p>5 баллов: построена линейная парная регрессия с заданной функцией потерь, по ней выполнен прогноз</p>	
--	--	--	--