1. **Паспорт образовательной программы**

**«Знакомство с Machine Learning»**

|  |  |
| --- | --- |
| **Версия программы** | 01 |
| **Дата версии** | 08.10.2020 г. |

|  |  |
| --- | --- |
| **Версия программы** | 01 |
| **Дата версии** | 08.10.2020 г. |

1. **Сведения о провайдере:**

|  |  |  |
| --- | --- | --- |
| **№** |  | **Содержание** |
| **1.1** | Провайдер | Общество с ограниченной ответственностью «Юниум» |
| **1.2** | Логотип образовательной организации |  |
| **1.3** | Провайдер, ИНН | 7720364967 |
| **1.4** | Ответственный за программу, ФИО | Кондратьев Дмитрий Сергеевич |
| **1.5** | Ответственный за программу, должность | Генеральный директор |
| **1.6** | Ответственный за программу, телефон | +7 (903) 503-46-46 |
| **1.7** | Ответственный за программу,  Е-mail | dk@unium.ru |

1. **Основные данные:**

|  |  |  |
| --- | --- | --- |
| **№** |  | **Содержание** |
| **2.1** | Название программы | «Знакомство с Machine Learning» |
| **2.2** | Ссылка на страницу программы | http://edu2035.info/ |
| **2.3** | Формат обучения | онлайн |
|  | Подтверждение от ОО наличия возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа | Образовательная организация подтверждает возможность реализации программы с применением электронного обучения и (или) дистанционных образовательных технологий. Для реализации образовательной программы используется платформа «GetCourse». |
| **2.4** | Уровень сложности | Начальный |
| **2.6** | Количество академических часов | 72 ак. часа |
|  | Практикоориентированный характер образовательной программы: не менее 50 % трудоёмкости учебной деятельности отведено практическим занятиям и (или) выполнению практических заданий в режиме самостоятельной работы (кол-во академических часов) | Образовательная программа имеет практикоориентированный характер, 48 ак. час трудоемкости образовательной программы отведено практическим занятиям и выполнению практических заданий в режиме самостоятельной работы. |
| **2.6** | Стоимость обучения одного обучающегося по образовательной программе, а также предоставление ссылок на 3 (три) аналогичные образовательные программы иных организаций, осуществляющих обучение, для оценки объективности стоимости или обоснование уникальности представленной образовательной программы в случае отсутствия аналогичных образовательных программ на рынке образовательных услуг | Стоимость обучения одного слушателя составляет  30 000 рублей.  Ссылки на аналогичные образовательные программы:  <https://netology.ru/programs/analytics-for-executives>  <https://newprolab.com/ru/bigdata/>  <https://www.bigdataschool.ru>  <https://bigdatateam.org/ru>  <https://www.cdo.training> |
| **2.7** | Минимальное количество человек на курсе | 100 |
| **2.8** | Максимальное количество человек на курсе | 2 000 |
| **2.9** | Данные о количестве слушателей, ранее успешно прошедших обучение по образовательной программе | 400 |
| **2.10** | Формы аттестации | Тестирование и состязание на Kaggle |
|  | Указание на область реализации компетенций цифровой экономики, к которой в большей степени относится образовательная программа, в соответствии с Перечнем областей | Большие данные |

1. **Аннотация программы:**

Образовательная программа по знакомству с Machine Learning является основой для дальнейшего освоения одной из самых высокооплачиваемых в мире профессий Data Scientist. Курс знакомит слушателей с основными инструментами машинного обучения (алгоритмы классификации, кластеризации, прогнозирования и др.) и анализа данных, вырабатывает практические навыки написания программного кода. Программа ориентирована на начальный уровень знакомства с ML, первые занятия посвящены основам применяемого при создании самообучающихся алгоритмов математического аппарата (некоторые разделы линейной алгебры и математического анализа). Каждое занятие сопровождается практическими примерами и заданиями. В рамках программы будет изучено свыше 15 различных математических методов и более 10 библиотек Python (numpy, scipy, MatplotLib и др.).

Участники, успешно завершившие курс, смогут самостоятельно писать (программировать) основные алгоритмы машинного обучения, а также работать с большими выборками данных (препроцессинг данных, отбор и формирование признаков, выбор оптимальных библиотек и их применение, выбор оптимальных метрик качества с учетом бизнес-задачи). Программа рассчитана на слушателей, имеющих базовые знания в области линейной алгебры, теории вероятности и вычислительной математики.

По результатам обучения у участников будет возможность поучаствовать в реальном состязании на Kaggle.

УТВЕРЖДАЮ

Генеральный директор

ООО «Юниум»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.С. Кондратьев

«8» октября 2020 г

1. ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА (ДПО)

Общество с ограниченной ответственностью «Юниум»

УТВЕРЖДАЮ

Генеральный директор

ООО «Юниум»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.С. Кондратьев

«8» октября 2020 г

**ОСНОВНАЯ ПРОФЕССИОНАЛЬНАЯ ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПОДГОТОВКИ**

**(повышение квалификации)**

**по направлению подготовки**

**«Знакомство с Machine Learning»**

72 ак. часа

Форма обучения: онлайн

г. Москва, 2020 г.

**ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

1. **Цель программы:**

Освоение теоретических основ и приобретение практических навыков в области разработки Machine Learning-проектов: от сбора и анализа данных до оптимизации моделей ML.

1. **Планируемые результаты обучения:** 
   1. **Знание (осведомленность в областях):**
      1. Знать базовые понятия математического аппарата;
      2. Знать основы математической статистики;
      3. Знать основы линейной алгебры и математического анализа;
      4. Знать разведочный анализ данных (EDA);
      5. Знать возможности библиотеки языка Python для работы с анализом данных и их визуализаций;
      6. Знать Индекс Джини, понятие энтропии и прироста информации по Шеннону;
      7. Знать основы моделей, построенных на композиции алгоритмов;
      8. Знать метод наименьших квадратов;
      9. Знать метод стохастического градиентного спуска;
      10. Знать метод максимального правдоподобия;
      11. Знать основы нормального и равномерного распределения и регуляции;
      12. Знать способ ML- обучение без учителя;
      13. Знать метод ближайших соседей, потенциалов и k-средних;
      14. Знать метод DBScan;
      15. Знать методы парзеновского окна;
      16. Знать теоретические основы градиентного бустинга;
   2. **Умение (способность к деятельности):**
      1. Уметь решать базовые статистические задачи, задачи линейной алгебры и математического анализа;
      2. Уметь с помощью возможностей языка Python анализировать данные;
      3. Уметь с помощью возможностей языка Python визуализировать данные;
      4. Уметь принимать решения в «узлах дерева»;
      5. Уметь решать задачи кластеризации в машинном обучении.
   3. **Навыки (использование конкретных инструментов):**
      1. Владеть навыками использования базовых библиотек pandas/numpy/matplotlib/seaborn/sklearn языка python;
      2. Владеть алгоритмом построения и использования модели машинного обучения «Деревья решений»;
      3. Владеть навыками построения модели простого и взвешенного голосования;
      4. Владеть навыками первоначального анализа данных, визуализации результатов анализа и оценки эффективности применимости различных моделей машинного обучения;
      5. Владеть навыками прогнозирования с использованием моделей машинного обучения;
      6. Владеть навыками решения практических задач машинного обучения.
2. **Категории слушателей:**
   1. Образование: среднее профессиональное и/или высшее образование
   2. Квалификация: рекомендовано базовое понимание математического аппарата и базовые навыки работы на языке Python.
   3. Наличие опыта профессиональной деятельности: не требуется.
   4. Предварительное освоение иных дисциплин/курсов/модулей: не требуется.
3. **Учебный план программы «Знакомство с Machine Learning»:**

В рабочем учебном плане отображается логическая последовательность освоения дисциплин, обеспечивающих формирование компетенций. Указывается общая трудоемкость дисциплин, а также их общая и аудиторная трудоемкость в часах.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Наименование**  **модулей** | **Всего часов** | **Виды учебных занятий** | | | **Форма контроля** |
| **Лекции** | **Практические занятия** | **Сам. работа** |
| 1 | Модуль 1. «Математические основы машинного обучения и базовые навыки работы с библиотеками анализа данных на языке Python» | 20 | 10 | 5 | 5 | Зачет |
| 2 | Модуль 2. «Базовые модели машинного обучения: задачи классификации и регрессии» | 12 | 6 | 3 | 3 | Зачет |
| 3 | Модуль 3. «Обучение без учителя» | 8 | 4 | 2 | 2 | Зачет |
| 4 | Модуль 4. «Основные подходы к практическим задачам» | 8 | 4 | 2 | 2 | Зачет |
| 5 | Модуль 5. «Применение полученных навыков , полученных в процессе обучения (соревнования на платформе Kaggle)» | 24 | 0 | 0 | 24 | Зачет |
|  | ИТОГО | 72 | 24 | 12 | 36 |  |

1. **Календарный план-график реализации образовательной «Знакомство с Machine Learning»:**

Дата начала обучения: 1 ноября 2020 года.

Дата завершения обучения: 22 ноября 2020 года.

Указания на периодичность набора групп: 1 группа в месяц.

|  |  |  |  |
| --- | --- | --- | --- |
| **№**  **п/п** | **Наименование**  **модулей** | **Трудоёмкость (ак. час)** | **Сроки обучения** |
|
| 1 | Модуль 1. «Математические основы машинного обучения и базовые навыки работы с библиотеками анализа данных на языке Python» | 20 | 01.11.2020 – 07.11.2020 |
| 2 | Модуль 2. «Базовые модели машинного обучения: задачи классификации и регрессии» | 12 | 08.11.2020 – 11.11.2020 |
| 3 | Модуль 3. «Обучение без учителя» | 8 | 12.11.2020 – 13.11.2020 |
| 4 | Модуль 4. «Основные подходы к практическим задачам» | 8 | 14.11.2020 – 16.11.2020 |
| 5 | Модуль 5, «Применение полученных навыков, полученных в процессе обучения (соревнования на платформе Kaggle)» | 24 | 17.11.2020 – 22.11.2020 |
|  | ИТОГО | **72** | **Зачёт** |

1. **Учебно-тематический план образовательной программы «Знакомство с Machine Learning»:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Наименование**  **модулей/тем** | **Всего часов** | **Виды учебных занятий** | | | **Форма контроля** |
| **Лекции** | **Практические занятия** | **Сам. работа** |
| **1** | **Модуль1. «Математические основы машинного обучения и базовые навыки работы с библиотеками анализа данных на языке Python»** | **20** | **10** | **5** | **5** | **Зачет** |
| 1.1 | Тема 1.1. Математические основы ML | 8 | 4 | 2 | 2 |  |
| 1.2 | Тема 2.2. Разведочный анализ данных (EDA) | 12 | 6 | 3 | 3 |  |
| **2** | **Модуль 2. «Базовые модели машинного обучения: задачи классификации и регрессии»** | **12** | **6** | **3** | **3** | **Зачет** |
| 2.1 | Тема 2.1. Деревья решений | 4 | 2 | 1 | 1 |  |
| 2.2 | Тема 2.2. Базовые композици алгоритмов | 4 | 2 | 1 | 1 |  |
| 2.3 | Тема 2.3. МНК и линейные модели | 4 | 2 | 1 | 1 |  |
| **3** | **Модуль 3. «Обучение без учителя»** | **8** | **4** | **2** | **2** | **Зачет** |
| 3.1 | Тема 3.1. Кластеризация | 4 | 2 | 1 | 1 |  |
| 3.2 | Тема 3.2. Метрические алгоритмы | 4 | 2 | 1 | 1 |  |
| 4 | **Модуль 4. «Основные подходы к практическим задачам»** | **8** | **4** | **2** | **2** | **Зачет** |
| 4.1 | Тема 4.1. Feature Encoding | 4 | 2 | 1 | 1 |  |
| 4.2 | Тема 4.2. Бустинг | 4 | 2 | 1 | 1 |  |
| 5. | **Модуль 5. Применение полученных навыков, полученных в процессе обучения (соревнования на платформе Kaggle)** | **24** | **0** | **0** | **24** |  |
| 5.1 | Практическая работа Соревнование на платформе Kaggle | 24 | 0 | 0 | 24 |  |
|  | **ИТОГО** | **72** | **24** | **12** | **36** | **Зачет** |

1. **Учебная (рабочая) программа повышения квалификации «Знакомство с Machine Learning»:**

**Модуль 1. «Математические основы машинного обучения и базовые навыки работы с библиотеками анализа данных на языке Python» (20 ак.ч.)**

Тема 1.1 «Математические основы» (8 ак.ч.)

Повторение базовых понятий математического аппарата, математическая статистика, линейная алгебра и математический анализ.

Тема 1.2 «Разведочный анализ данных (EDA)» (12 ак.ч.)

Изучение базовых и углубленных возможностей библиотек языка Python для работы с анализом данных и их визуализаций.

**Модуль 2. «Базовые модели машинного обучения: задачи классификации и регрессии» (12 ак.ч.)**

Тема 2.1 «Деревья решений» (4 ак.ч.)

Пояснение основ алгоритма построения модели машинного обучения под названием «Деревья решений». Индекс Джини, понятие энтропии и прироста информации по Шеннону. Принятие решения в узлах дерева.

Тема 2.2 «Базовые композиция алгоритмов» (4 ак.ч.)

Основы моделей, построенных на композиции алгоритмов. Модель простого и взвешенного голосования.

Тема 2.3 «МНК и линейные модели» (4 ак.ч.)

Метод наименьших квадратов. Линейная регрессия. Градиентный спуск. Метод максимального правдоподобия. Нормальное и равномерное распределение. Регуляризация.

**Модуль 3. «Обучение без учителя» (8 ак.ч.)**

Тема 3.1 «Кластеризация» (4 ак.ч.)

Задача кластеризации в машинном обучении. Основные подходы к обучению без учителя. Метод ближайших соседей. Метрики кластеризации. Метод главных компонент. Агломеративная кластеризация.

Тема 3.2 «Метрические алгоритмы» (4 ак.ч.)

Метод ближайших соседей. Метод потенциалов. Метод k-средних. DBScan. Методы парзеновского окна.

**Модуль 4. «Основные подходы к практическим задачам» (8 ак.ч.)**

Тема 4.1 «Feature Encoding» (4 ак.ч.)

One-Hot кодирование. Мешок слов. Kernel trick. Кодирование средним. Базовые методы работы с текстовыми данными.

Тема 4.2 «Бустинг» (4 ак.ч.)

Теоретические основы градиентного бустинга. Современный взгляд на бустинг. Adaboost. Переобучение. Функции потерь.

**Модуль 5. «Применение полученных навыков, полученных в процессе обучения (соревнования на платформе Kaggle)» (24 ак.ч.)**

Тема 5.1 «Соревнование» (24 ак.ч.)

Применение практических навыков, полученных в процессе обучения, в формате соревнования выпускников.

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| **1** | **Модуль 1. «Математические основы машинного обучения и базовые навыки работы с библиотеками анализа данных на языке Python»** | Практическое задание 1 | Тесты по пройденным темам лекций. Решение математических заданий с использование полученных теоретических и практических навыков. Написание кода на языке Python с использованием библиотек анализа данных и визуализации. |
|  | **Модуль 2. «Базовые модели машинного обучения: задачи классификации и регрессии»** | Практическое задание 2 | Программная реализация одного из методов классификации на языке Python. Сравнение реализации обучаемого с существующими решениями. |
|  | **Модуль 3. «Обучение без учителя»** | Практическое задание 3 | Программная реализация одного из методов обучения без учителя на языке Python. Сравнение реализации обучаемого с существующими решениями. |
|  | **Модуль 4. «Основные подходы к практическим задачам»** | Практическое задание 4 | Обработка текстовых данных с использованием реализаций моделей, написанных во 2 и 3 модулях. |
|  | **Модуль 5. «Применение полученных навыков, полученных в процессе обучения (соревнования на платформе Kaggle)»** | Практическое задание 5 | Решение практической задачи в формате соревнования выпускников между собой. |

**8.Оценочные материалы по образовательной программе**

**8.1. Вопросы тестирования по модулям**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ модуля** | **Вопросы входного тестирования** | **Вопросы промежуточного тестирования** | **Вопросы итогового тестирования** |
| **1** | Перечислите основные эконометрические методы анализа данных. Найдите обратную матрицу А, и перемножьте ее на матрицу B. Возможно ли в данном случае перемножение? | Назовите примеры алгоритмов машинного обучения. Опишите алгоритм наиболее правильной со статистической точки зрения работы с данными по шагам. | Загрузите датасет по смертности пассажиров “Титаника”, постройте для него гистограмму, очистите от выбросов, проведите регрессионный анализ (влияние уровня доходов пассажиров на их смертность). Представьте все необходимые графики. |
| **2** | Объясните, в чем заключается разница между линейной и логарифмической регрессиями. | Опишите принцип работы алгоритма классификации данных. | Загрузите датасет по смертности пассажиров “Титаника” и выявите признаки, влияющие на смертность (1 - пассажир погиб, 0 - пассажир остался живым). |
| **3** | Как Вы понимаете понятие “кластер”? | Опишите преимущества и недостатки нейронной сети Кохонена. | Постройте на основе загруженного датасета по Титанику сеть радиально-базисных функций (RBF). |
| **4** | Назовите наиболее встречаемые в практике алгоритмы машинного обучения. | На основе исторического курса пары рубль/доллар за 2015, 2016, 2017, 2018, 2019, 1 половины 2020 года предскажите значение курса рубля на сентябрь 2020 года. Соответствуют ли Ваши прогнозные результаты реальным? Какие факторы Вы не учли? Что можно добавить в Ваш анализ для повышения точности прогнозирования? | На основе открытых сведений об энергопотреблении зданий в Нью-Йорке предскажите рейтинг энергопотребления (Energy Star Score) здания и выявите наиболее влиятельные признаки. Постройте Pair Plots of Energy Data. |
| **5** | Назовите иерархию “званий” в Kaggle. Каковы критерии для достижения каждого из званий? | Решите настоящее задание про Титаник в соревновании Kaggle (<https://www.kaggle.com/c/titanic>). Что удалось и что не получилось? Как Вы думаете, почему? | Финальным заданием курса будет соревнование в Kaggle. Его условия будут зависеть от скорости и успешности прохождения слушателями материалов курса, а также уровня компетенции слушателей. |

**8.2.** О**писание показателей и критериев оценивания, шкалы оценивания.**

**Критерии оценивания тестирования**:

«Зачёт» - получают обучающиеся, справившиеся с тестом 65-100% (10 – 15 верных ответов);

«Незачёт» - соответствует работе, содержащей менее 65% правильных ответов (1- 9 верных ответов).

**Критерии оценивания выполнения практического задания:**

«Зачёт» - получают обучающиеся, успешно справившиеся с заданием: варианты оценки - “полностью справился с заданием”, “справился с заданием в основном, цель достигнута, но есть несущественные недочеты”;

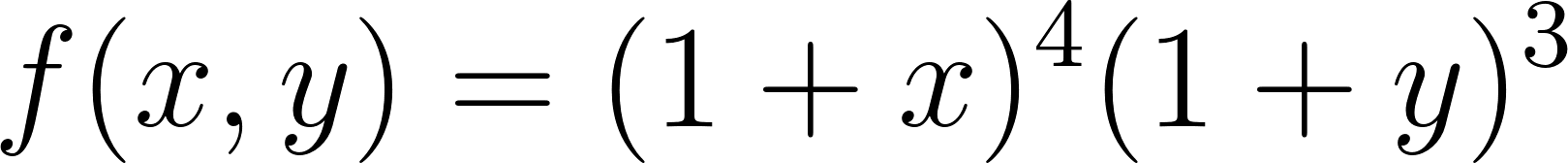
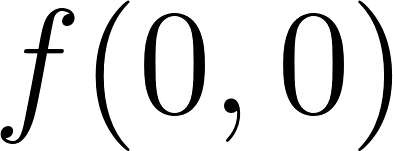
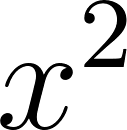
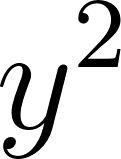
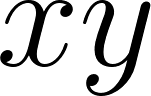
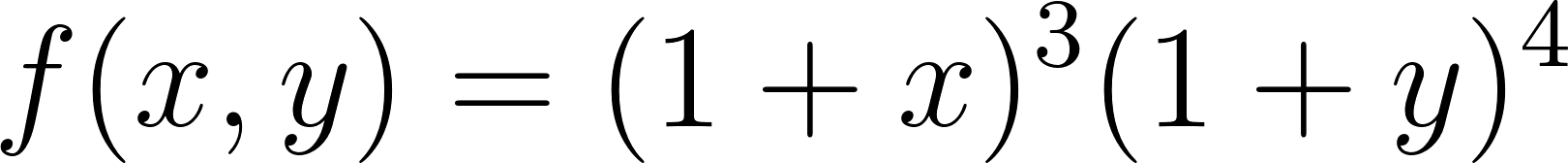
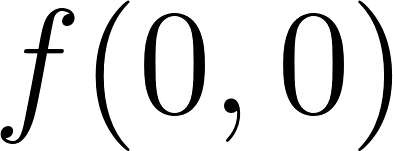
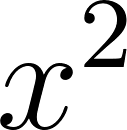
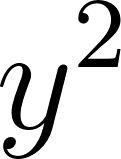
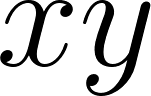
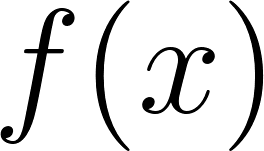
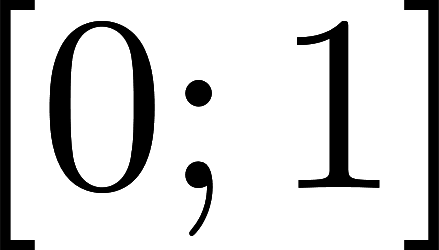
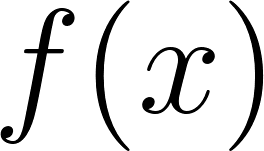
«Незачёт» - ставится в случае невыполнения задания: варианты оценки - “не справился с заданием совсем”, “справился с заданием частично, но целевые результаты не достигнуты”.

**8.3. Примеры контрольных заданий по модулям или всей образовательной программе**

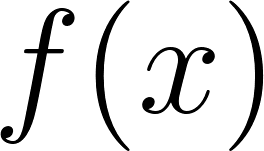
Пример практической задачи: “Строительная фирма представила на городской конкурс два проекта. Руководство фирмы считает, что вероятность выигрыша только первого проекта равна 0.2, только второго — 0.5, обоих проектов — 0.3. Чему равна вероятность того, что выиграет хотя бы один проект?”

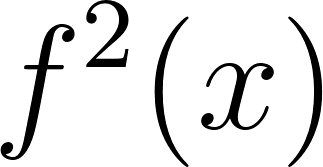
**8.4.** Т**есты и обучающие задачи (кейсы), иные практико-ориентированные формы заданий.**

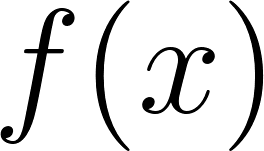
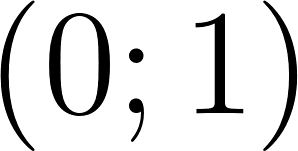
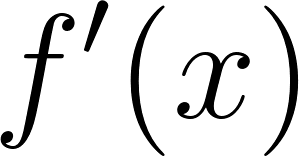
Тесты для входного и выходного тестирования:

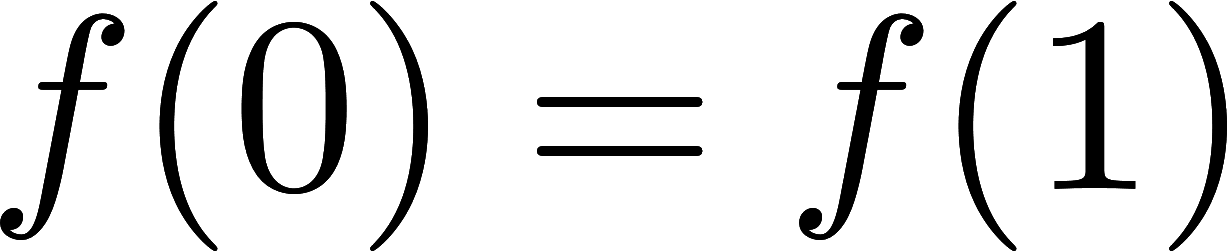
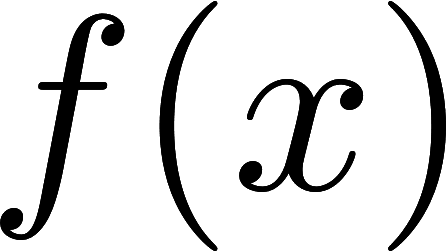
1. Разложить функцию [](https://www.codecogs.com/eqnedit.php?latex=f(x%2C%20y)%3D(1%2Bx)%5E4(1%2By)%5E3%250) в ряд Тейлора в точке [](https://www.codecogs.com/eqnedit.php?latex=f(0%2C%200)%250) и найти сумму коэффициентов перед [](https://www.codecogs.com/eqnedit.php?latex=x%250), [](https://www.codecogs.com/eqnedit.php?latex=x%5E2%250), [](https://www.codecogs.com/eqnedit.php?latex=y%250), [](https://www.codecogs.com/eqnedit.php?latex=y%5E2%250) и [](https://www.codecogs.com/eqnedit.php?latex=xy%250). Ответ округлите до 2 знаков после запятой.  
   **Ответ: 28**
2. Разложить функцию [](https://www.codecogs.com/eqnedit.php?latex=f(x%2C%20y)%3D(1%2Bx)%5E3(1%2By)%5E4%250) в ряд Тейлора в точке [](https://www.codecogs.com/eqnedit.php?latex=f(0%2C%200)%250) и найти сумму коэффициентов перед [](https://www.codecogs.com/eqnedit.php?latex=x%250), [](https://www.codecogs.com/eqnedit.php?latex=x%5E2%250), [](https://www.codecogs.com/eqnedit.php?latex=y%250), [](https://www.codecogs.com/eqnedit.php?latex=y%5E2%250) и [](https://www.codecogs.com/eqnedit.php?latex=xy%250). Ответ округлите до 2 знаков после запятой.  
   **Ответ: 28**
3. Функция [](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) определена и непрерывна на отрезке [](http://www.texrendr.com/?eqn=%5B0%3B%201%5D%250), при этом [](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) достигает наибольшего значения ровно в двух точках, не совпадающих ни с 0, ни с 1. Тогда

[Одиночный выбор]

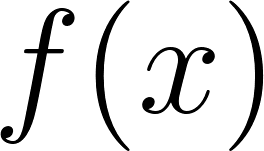
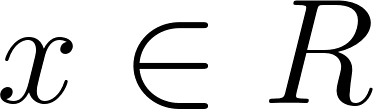
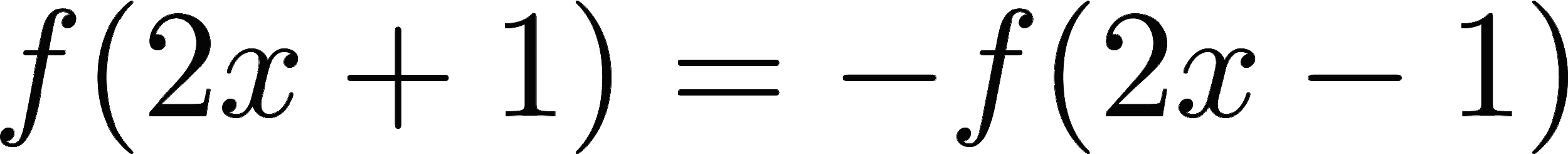
· функция [](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) достигает наименьшего значения не менее, чем в трех точках.

· функция [](https://www.codecogs.com/eqnedit.php?latex=f%5E2(x)%250) достигает наибольшего значения не менее, чем в двух точках.

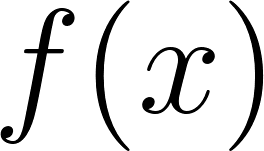
· **если функция** [****](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) **дифференцируема на интервале** [****](https://www.codecogs.com/eqnedit.php?latex=(0%3B%201)%250)**, то производная** [****](https://www.codecogs.com/eqnedit.php?latex=f'(x)%250) **равна нулю не менее, чем в трех точках.**

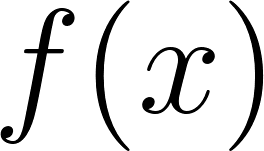
· если [](http://www.texrendr.com/?eqn=f(0)%3Df(1)%250), то функция [](http://www.texrendr.com/?eqn=f(x)%250) достигает наименьшего значения не менее чем в трех точках

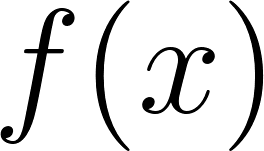
все утверждения выше ложные

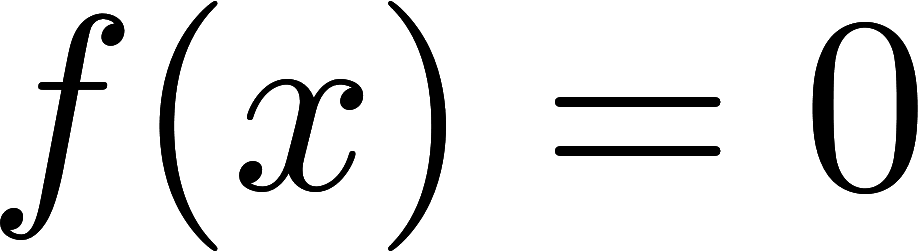
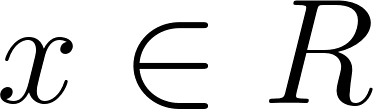
1. Известно, что для непрерывной функции [](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) при всех [](https://www.codecogs.com/eqnedit.php?latex=x%20%5Cin%20R%250) выполнено равенство [](http://www.texrendr.com/?eqn=f(2x%2B1)%3D-f(2x-1)%250). Найдите **ложное** утверждение:

[Одиночный выбор]

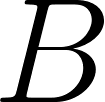
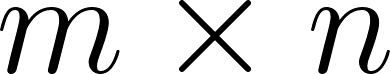
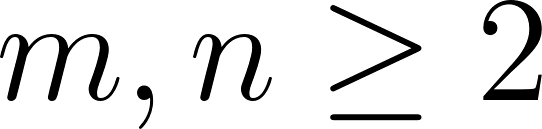
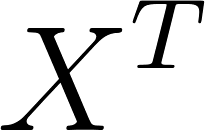
· функция [](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) ограниченная.

· функция [](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) периодическая.

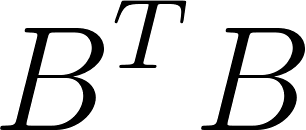
· **функция** [****](https://www.codecogs.com/eqnedit.php?latex=f(x)%250) **нечетная.**

· [](http://www.texrendr.com/?eqn=f(x)%20%3D%200%250) при некотором [](https://www.codecogs.com/eqnedit.php?latex=x%20%5Cin%20R%250)

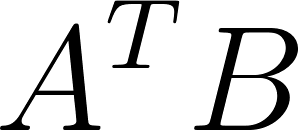
· среди утверждений выше есть ложное

1. Даны матрицы [](https://www.codecogs.com/eqnedit.php?latex=A%250) и [](https://www.codecogs.com/eqnedit.php?latex=B%250) размера [](https://www.codecogs.com/eqnedit.php?latex=m%20%5Ctimes%20n%250), где [](https://www.codecogs.com/eqnedit.php?latex=m%2Cn%20%5Cge%202%250), у которых строки линейно независимы. Через [](https://www.codecogs.com/eqnedit.php?latex=X%5ET%250) обозначим матрицу, транспонированную к матрицу [](https://www.codecogs.com/eqnedit.php?latex=X%250). Тогда:

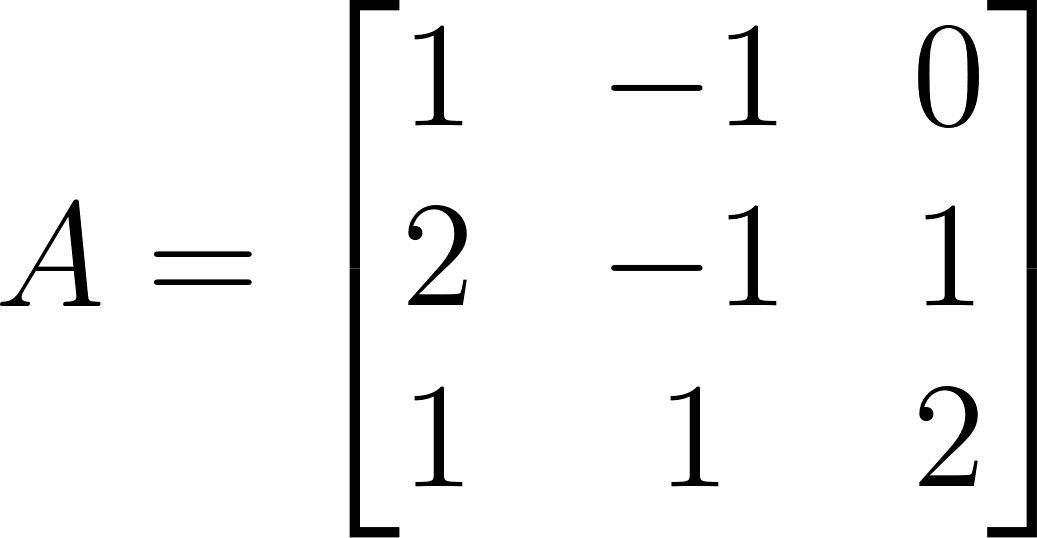
· **матрица** [****](https://www.codecogs.com/eqnedit.php?latex=A%20A%5ET%250) **невырождена**

· матрица [](https://www.codecogs.com/eqnedit.php?latex=B%5ETB%250) невырожденная

· матрица [](https://www.codecogs.com/eqnedit.php?latex=AB%5ET%250) невырожденная

· матрица [](https://www.codecogs.com/eqnedit.php?latex=A%5ETB%250) невырожденная

· все утверждения данные выше неверные

1. Задана матрица [](https://www.codecogs.com/eqnedit.php?latex=A%3D%5Cbegin%7Bbmatrix%7D1%20%26%20-1%20%26%200%20%5C%5C%202%20%26%20-1%20%26%201%20%5C%5C%201%20%26%201%20%26%202%5Cend%7Bbmatrix%7D%250). Тогда:

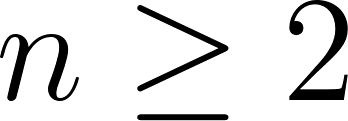
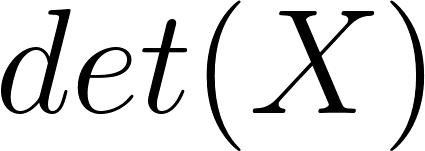
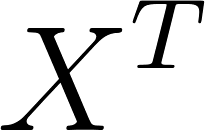
· матрица [](https://www.codecogs.com/eqnedit.php?latex=A%250) имеет 3 различных вещественных собственных числа

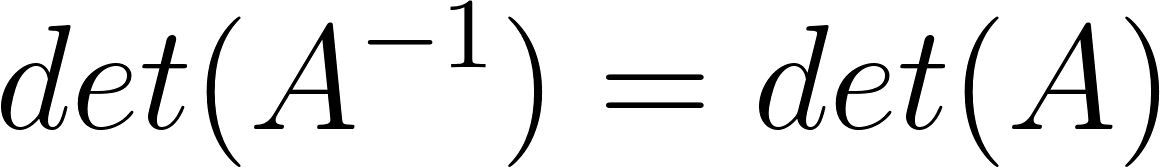
· у матрицы [](https://www.codecogs.com/eqnedit.php?latex=A%250) для всех вещественных собственных чисел их геометрическая кратность совпадает с алгебраической

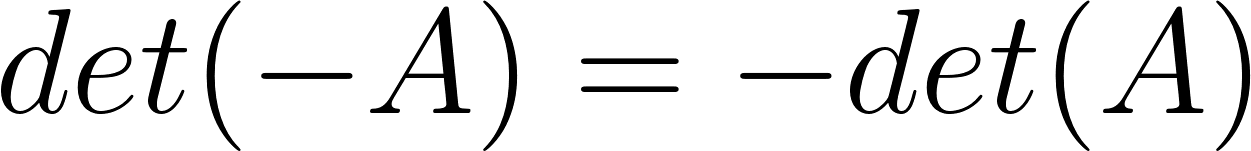
· ноль является собственным числом матрицы [](https://www.codecogs.com/eqnedit.php?latex=A%250), и размерность соответствующего собственного подпространства равна 2

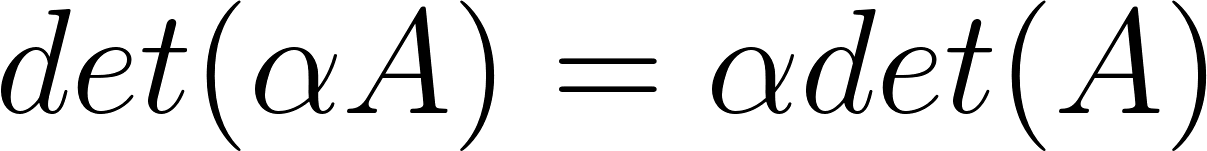
· **наибольшее вещественное собственное число матрицы** [****](https://www.codecogs.com/eqnedit.php?latex=A%250) **равно 2**

· все утверждения данные выше неверные

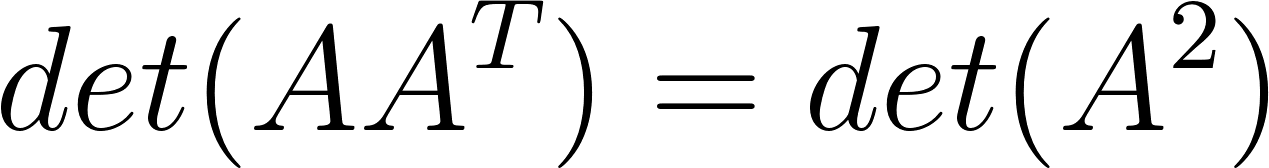
1. Дана квадратичная матрица [](https://www.codecogs.com/eqnedit.php?latex=A%250) порядка [](https://www.codecogs.com/eqnedit.php?latex=%20n%20%5Cge%202%250). Через [](https://www.codecogs.com/eqnedit.php?latex=det(X)%250) обозначим определитель любой квадратной матрицы [](https://www.codecogs.com/eqnedit.php?latex=X%250), а через [](https://www.codecogs.com/eqnedit.php?latex=X%5ET%250) обозначим матрицу, транспонированную к матрице [](https://www.codecogs.com/eqnedit.php?latex=X%250). Тогда:

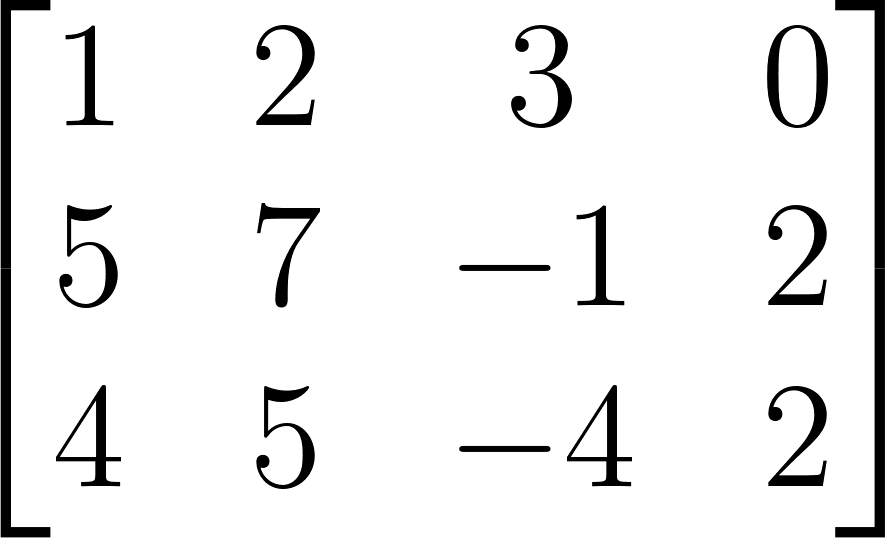
· [](https://www.codecogs.com/eqnedit.php?latex=det(A%5E%7B-1%7D)%3Ddet(A)%250)

· [](https://www.codecogs.com/eqnedit.php?latex=det(-A)%3D-det(A)%250)

· [****](https://www.codecogs.com/eqnedit.php?latex=det(%5Calpha%20A)%3D%5Calpha%20det(A)%250)

· [](https://www.codecogs.com/eqnedit.php?latex=det(A%2BB)%3Ddet(A)%2Bdet(B)%250)

· [](https://www.codecogs.com/eqnedit.php?latex=det(AA%5ET)%3Ddet(A%5E2)%250)**(ответ)**

1. Ранг матрицы [](https://www.codecogs.com/eqnedit.php?latex=%5Cbegin%7Bbmatrix%7D1%20%26%202%20%26%203%20%26%200%20%5C%5C%205%20%26%207%20%26%20-1%20%26%202%20%5C%5C%204%20%26%205%20%26%20-4%20%26%202%5Cend%7Bbmatrix%7D%250) равен:

[Одиночный выбор]

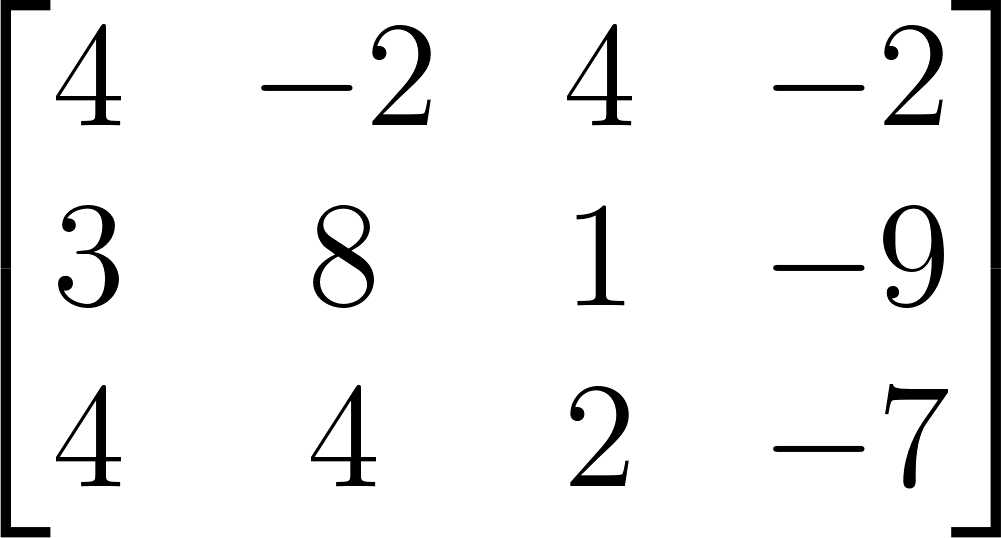
· 0

· 1

· **2**

· 3

· 4

1. Ранг матрицы [](https://www.codecogs.com/eqnedit.php?latex=%5Cbegin%7Bbmatrix%7D4%20%26%20-2%20%26%204%20%26%20-2%20%5C%5C%203%20%26%208%20%26%201%20%26%20-9%20%5C%5C%204%20%26%204%20%26%202%20%26%20-7%5Cend%7Bbmatrix%7D%250) равен:

[Одиночный выбор]

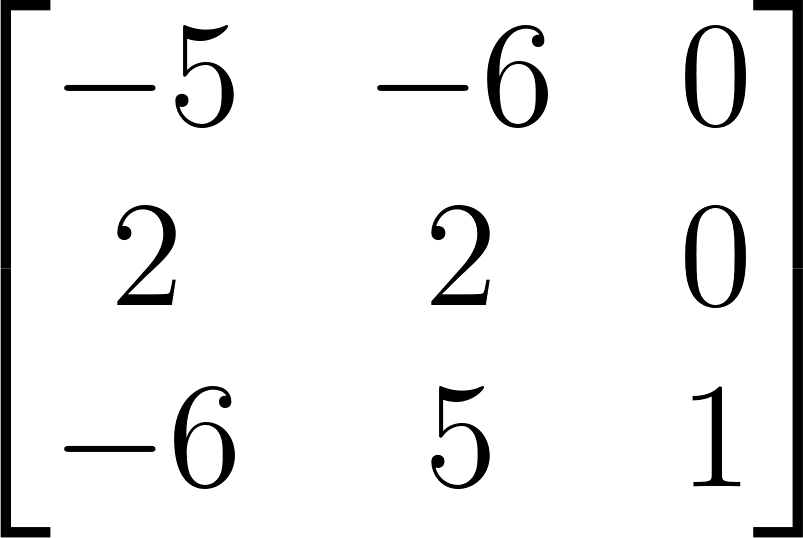
· 0

· 1

· 2

· **3**

· 4

1. Определитель матрицы [](https://www.codecogs.com/eqnedit.php?latex=%5Cbegin%7Bbmatrix%7D-5%20%26%20-6%20%26%200%20%5C%5C%202%20%26%202%20%26%200%20%5C%5C%20-6%20%26%205%20%26%201%5Cend%7Bbmatrix%7D%250) равен:

[Одиночный выбор]

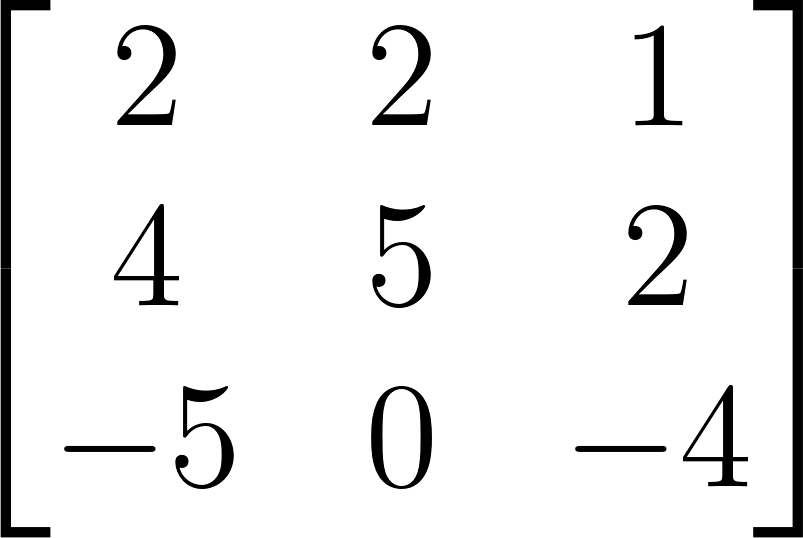
· -1

· 0

· 1

· **2**

· не существует

1. Определитель матрицы [](https://www.codecogs.com/eqnedit.php?latex=%5Cbegin%7Bbmatrix%7D2%20%26%202%20%26%201%20%5C%5C%204%20%26%205%20%26%202%20%5C%5C%20-5%20%26%200%20%26%20-4%5Cend%7Bbmatrix%7D%250) равен:

[Одиночный выбор]

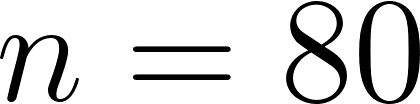
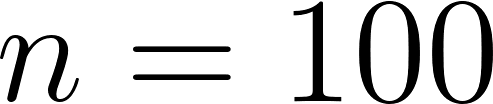
· 1

· 3

· -1

· **-3**

· не существует

1. Восемь процентов взрослого населения большого города являются безработными. Найдите дисперсию в случайной выборке объёма [](https://www.codecogs.com/eqnedit.php?latex=n%3D80%250). Ответ округлите до 2 знаков после запятой.  
   [Одиночный выбор]  
   · **5.89**· 10.45  
   · 7.28  
   · 9.36  
   · 11.62
2. Два процента взрослого населения большого города являются безработными. Найдите дисперсию в случайной выборке объёма [](https://www.codecogs.com/eqnedit.php?latex=n%3D100%250). Ответ округлите до 2 знаков после запятой.  
   [Одиночный выбор]  
   · 5.89  
   · 5.64  
   · **1.96**· 4.05  
   · 9.18
3. В лотерее на 1% билетов выпадает выигрыш в 200 рублей, на 0.01% билетов выпадает выигрыш в 1000 рублей, а остальные билеты без выигрыша. Найдите средний выигрыш в этой лотерее (в расчете на один билет). Ответ округлите до 2 знаков после запятой.  
   [Одиночный выбор]  
   · 2.68  
   · 6.91  
   · 2.76  
   · **2.10**· 8.19
4. В лотерее на 2% билетов выпадает выигрыш в 300 рублей, на 0.05% билетов выпадает выигрыш в 500 рублей, а остальные билеты без выигрыша. Найдите средний выигрыш в этой лотерее (в расчете на один билет). Ответ округлите до 2 знаков после запятой.  
   [Одиночный выбор]  
   · 8.00  
   · **6.25**· 0.85  
   · 2.10  
   · 6.70
5. Среднее число самолетов, взлетающих с аэродрома за сутки, равно 10. С помощью распределения Пуассона найти вероятность того, что за 6 часов взлетит не менее двух самолетов. Ответ округлите до 3 знаков после запятой.  
   **Ответ: 0.713**
6. Среднее число самолетов, взлетающих с аэродрома за сутки, равно 10. С помощью распределения Пуассона найти вероятность того, что за 12 часов взлетит не менее трех самолетов. Ответ округлите до 3 знаков после запятой.  
   **Ответ: 0.560**
7. Строительная фирма представила на городской конкурс два проекта. Руководство фирмы считает, что вероятность выигрыша первого проекта равна 0.5, второго — 0.4, обоих проектов — 0.2.  
   Чему равна вероятность того, что выиграет хотя бы один проект?  
   [Одиночный выбор]  
   · 0.6  
   · **0.7**· 0.5  
   · 0.3  
   · Ни один из вариантов выше не содержит правильного ответа
8. Строительная фирма представила на городской конкурс два проекта. Руководство фирмы считает, что вероятность выигрыша первого проекта равна 0.4, второго — 0.4, обоих проектов — 0.2.  
   Чему равна вероятность того, что выиграет хотя бы один проект?  
   [Одиночный выбор]  
   · **0.6**  
   · 0.7· 0.5  
   · 0.3  
   · и один из вариантов выше не содержит правильного ответа
9. Строительная фирма представила на городской конкурс два проекта. Руководство фирмы считает, что вероятность выигрыша первого проекта равна 0.4, второго — 0.4, обоих проектов — 0.2.  
   Чему равна вероятность того, что первый проект выиграет конкурс, а второй нет?  
   [Одиночный выбор]  
   · **0.24**  
   · 0.3· 0.4  
   · 0.16  
   · Ни один из вариантов выше не содержит правильного ответа
10. Экзамен формата «Множественный выбор» содержит 6 вопросов, на каждый из которых дается 5 вариантов ответа, и среди них только один правильный. Для успешной сдачи экзамена надо ответить правильно не менее, чем на 4 вопроса из 6. Предположим, что студент не знает правильного ответа ни на один вопрос и выбирает стратегию чистого угадывания. Чему равна вероятность того, что он сдаст экзамен? Ответ округлите до 3 знаков после запятой.  
    [Одиночный выбор]  
    · **0.017**· 0.021  
    · 0.032  
    · 0.045  
    0.052
11. Вы купили облигацию, которая может оказаться "хорошей" с вероятностью 0.8 или "плохой" с вероятностью 0.2. "Хорошая" облигация с вероятностью 0.9 через год выплатит 100, а с вероятностью 0.1 не выплатит ничего. "Плохая" облигация через год с вероятностью 0.5 выплатит 100, а с вероятностью 0.5 не выплатит ничего. Через год облигация выплатила 100. Чему равна вероятность, что это "хорошая" облигация? Ответ округлите до 3 знаков после запятой.  
    **Ответ: 0.878**
12. Вы купили облигацию, которая может оказаться "хорошей" с вероятностью 0.7 или "плохой" с вероятностью 0.3. "Хорошая" облигация с вероятностью 0.8 через год выплатит 100, а с вероятностью 0.2 не выплатит ничего. "Плохая" облигация через год с вероятностью 0.6 выплатит 100, а с вероятностью 0.4 не выплатит ничего. Через год облигация выплатила 100. Чему равна вероятность, что это "хорошая" облигация? Ответ округлите до 3 знаков после запятой.  
    **Ответ: 0.757**

Тесты для выходного тестирования:

1. Какая библиотека на языке Python отвечает за быструю работу с массивами:  
   [Одиночный выбор]  
   · **Numpy**· Pandas  
   · Scipy  
   · Sklearn  
   · Seaborn
2. Какая библиотека на языке Python отвечает за удобную работу с табличными данными:  
   [Одиночный выбор]  
   · Numpy· **Pandas**  
   · Scipy  
   · Sklearn  
   · Seaborn
3. Какая библиотека на языке Python отвечает за набор из большинства стандартных моделей машинного обучения:  
   [Одиночный выбор]  
   · Numpy· Pandas  
   · Scipy  
   · **Sklearn**  
   · Seaborn
4. С помощью какого метода библиотеки Pandas можно вывести первые 5 строк DataFrame:  
   [Одиночный выбор]  
   · tail()· **head()**  
   · first\_5\_rows()  
   · show\_rows()
5. С помощью какого метода библиотеки Pandas можно вывести последние 5 строк DataFrame:  
   [Одиночный выбор]  
   · **tail()**· head()  
   · last\_5\_rows()  
   · last()
6. С помощью какого метода библиотеки Numpy можно создать матрицу из единиц размера 10 на 10:  
   [Одиночный выбор]  
   · zeros((10,10))· array((10,10))  
   · identity\_array((10,10))  
   · **ones((10,10))**
7. С помощью какого метода библиотеки Numpy можно создать матрицу из нулей размера 10 на 10:  
   [Одиночный выбор]  
   · **zeros((10,10)**)· array((10,10))  
   · identity\_array((10,10))  
   · ones((10,10))
8. С помощью какого метода библиотеки Numpy можно создать единичную матрицу размера 10 на 10:  
   [Одиночный выбор]  
   · zeros((10,10))· **eye((10,10))**  
   · identity\_array((10,10))  
   · ones((10,10))

**8.5.** О**писание процедуры оценивания результатов обучения.**

Результаты выполнения практического задания оцениваются по достижению либо недостижению основной цели задания. Зачет- задание выполнено, цель достигнута в основном, н незачет- задание не выполнено, цель не достигнута.

Результаты тестирования, а также ответов на вопросы оцениваются по проценту верных ответов: Зачет – более 65%, незачет – менее 65%.

**9.Организационно-педагогические условия реализации программы**

**9.1. Кадровое обеспечение программы**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Фамилия, имя, отчество (при наличии)** | **Место основной работы и должность, ученая степень и ученое звание (при наличии)** | **Ссылки на веб-страницы с портфолио (при наличии)** | **Фото в формате jpeg** | | **Отметка о полученном согласии на обработку персональных данных** |
| **1** | **Лоскутов Антон Алексеевич** | Текущее место работы:  Mail.ru Group  Программист-исследователь:  ● Создал с нуля системы антифрода и, на первых этапах,  занимался продвижением на рынок;  ● Разрабатываю и поддерживаю архитектуру проекта;  ● Организовал пайплайн работы команды;  ● Создал систему прогнозирования платежеспособности и LTV  пользователей;  ● Участвую в создании рекомендательной системы товаров.  Otus. Онлайн-образование  Преподаватель:  ● Математика для Data Science. Базовый курс;  ● Математика для Data Science. Продвинутый курс;  ● Machine Learning;  ● Advanced Machine Learning.  Образование:  МГУ им. М.В. Ломоносова, Механико-Математический факультет. Кафедра математических и  компьютерных методов анализа. Специалитет.  Дополнительные курсы:  SmartDec — Основы статического анализа программ.  Техносфера@Mail.ru - Анализ больших данных.  Сейчас являюсь старшим ментором.  Stanford (online) - CS231n Convolutional  Neural Network for  Visual Recognition. | GitHub: @LoskutovAnton.  Старший лектор открытого  курса по Machine Learning в  Mail.ru Group;  Лектор авторского открытого  курса по Machine Learning в  Mail.ru Group. |  | **Получено** | |
| **2** | **Мажаров Иван Андреевич** | Текущее место работы: программист-исследователь в отделе предиктивной аналитики Mail.Ru. Система антифрода, предсказание LTV, построение ML архитектуры, скоринг пользователей.  Образование: МГУ имени М.В. Ломоносова. Факультет ВМК. Бакалавриат. Специальность «Прикладная математика и информатика»;  МГУ имени М.В. Ломоносова. Факультет ВМК. Магистратура.  Специальность «Интеллектуальные системы».  Курсы:  Техносфера@mail.ru  Многопоточное программирование на языке С/С++,  Информационный поиск, Нейронные сети в машинном обучении,  Разработка веб-сервисов на Golang.  Stanford (online) CS224n: Natural Language Processing with Deep  Learning | Github: @IvanMaz.  Научная работа: Mazharov, Dobrov. Named Entity Recognition for Information Security  Domain. Selected Papers of the XX International Conference on Data  Analytics and Management in Data Intensive Domains (DAMDID/RCDL  2018). |  | **Получено** | |

**9.2. Учебно-методическое обеспечение и информационное сопровождение**

|  |  |
| --- | --- |
| **Учебно-методические материалы** | |
| Методы, формы и технологии | Методические разработки,  материалы курса, учебная литература |
| <https://www.anaconda.com/> | 1. Анализ больших наборов данных / Юре Лесковец, Ананди Раджараман, Джеффри Д. Ульман; пер. с англ. А.А. Слинкина. – Москва: ДМК Пресс, 2016. – 498 с. – Режим доступа: <http://znanium.com/catalog/product/1027845>  2. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер (пер. с англ. Инны Гайдюк) изд. М.: Манн, Иванов и Фербер, 2014.  3. Убийственные большие данные / Кэти О’Нил (пер. В. Дягтерева) изд.: АСТ, 2017.  4. Теоретический минимум по Big Data / Анналин Ын, Кеннет Су (пер. А.В. Тимохин) изд.: Питер, 2019.  5. BIG DATA. Вся технология в одной книге. Андреас Вайгенд, 2018.  6. Основы Data Science и Big Data / Дэви Силен, Арно Мейсман, Мохамед Али, изд. Питер, 2018.  7. Укрощение больших данных / Билл Фрэнкс (пер. А. Баранов), изд.: Манн, Иванов и Фербер, 2014.  Big data простым языком / Алексей Благирев, Наталья Хапаева, изд. АСТ, 2019. |

|  |  |
| --- | --- |
| **Информационное сопровождение** | |
| **Электронные**  **образовательные ресурсы** | **Электронные**  **информационные ресурсы** |
| https://www.kaggle.com | https://habr.com/ru/hub/machine\_learning/ |

**9.3. Материально-технические условия реализации программы**

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования,  программного обеспечения |
| Лекции | Zoom |
| Самостоятельные работы | Google Forms |
| Практические занятия | Kaggle.com |

**Приложение 1**

1. ПАСПОРТ КОМПЕТЕНЦИИ

ОСНОВНАЯ ПРОФЕССИОНАЛЬНАЯ ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПОДГОТОВКИ

(повышение квалификации)

по направлению подготовки

**«Знакомство с Machine Learning»**

ООО «ЮНИУМ»

УТВЕРЖДАЮ

Генеральный директор

ООО «Юниум»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.С. Кондратьев

«8» октября 2020 г

|  |  |  |
| --- | --- | --- |
| **1.** | **Наименование компетенции** | **Умение решать основные практические задачи машинного обучения** |
| 2. | Указание типа компетенции | Профессиональная |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | Формирование знаний и навыков по решению основных задач машинного обучения, включая задачи классификации, кластеризации и прогнозирования. |
| **4.** | **Дескриптор знаний, умений и навыков по уровням** | **Индикаторы** |
| Начальный уровень  (Компетенция недостаточно развита. Частично проявляет навыки, входящие в состав компетенции. Пытается, стремится проявлять нужные навыки, понимает их необходимость, но у него не всегда получается.) | Знать основы машинного обучения;  Знать особенности решения задач машинного обучения;  Знать базовые понятия математического аппарата;  Знать основы математической статистики;  Знать основы линейной алгебры и математический анализ;  Уметь решать базовые статистические задачи, а также задачи линейной алгебры и математического анализа;  Владеть навыками использования базовых библиотек pandas/numpy/matplotlib/seaborn/sklearn языка python; |
| Базовый уровень  (Уверенно владеет навыками, способен, проявлять соответствующие навыки в ситуациях с элементами неопределённости, сложности.) | Знать разведочный анализ данных (EDA);  Знать возможности библиотеки языка Python для работы с анализом данных и их визуализаций;  Знать Индекс Джини, понятие энтропии и прироста информации по Шеннону;  Знать основы моделей, построенных на композиции алгоритмов;  Уметь с помощью возможностей языка Python анализировать данные;  Уметь с помощью возможностей языка Python визуализировать данные;  Владеть навыками первоначального анализа данных, визуализации результатов анализа и оценки эффективности применимости различных моделей машинного обучения; |
| Продвинутый  (Владеет сложными навыками, способен активно влиять на происходящее, проявлять соответствующие навыки в ситуациях повышенной сложности.) | Знать метод наименьших квадратов;  Знать метод стохастического градиентного спуска;  Знать метод максимального правдоподобия;  Знать основы нормального и равномерного распределения и регуляции;  Знать способ машинного обучения –“обучение без учителя”;  Знать метод ближайших соседей, потенциалов и k-средних;  Уметь принимать решения в «узлах дерева»;  Уметь решать задачи кластеризации в машинном обучении;  Владеть алгоритмом использования построения модели машинного обучения «Деревья решений»;  Владеть навыками построения модели простого и взвешенного голосования. |
| Профессиональный  (Владеет сложными навыками, создает новые решения для сложных проблем со многими взаимодействую-щими факторами, предлагает новые идеи и процессы, способен активно влиять на происходящее, проявлять соответствующие навыки  в ситуациях повышенной сложности.) | Знать метод DBScan;  Знать методы парзеновского окна;  Знать теоретические основы градиентного бустинга.  Уметь решать различные типы задач машинного обучения;  Владеть навыками прогнозирования с использованием моделей машинного обучения;  Владеть навыками решения практических задач машинного обучения. |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | Математика, теория вероятностей, программирование |
| 6. | Средства и технологии оценки | Практические задания  Тесты |

**VI. Рекомендаций к программе от работодателей**: наличие не менее двух писем и/или подтверждения на цифровой платформе Государственной системы предоставления ПЦС от работодателей о рекомендации образовательной программы для реализации в рамках Государственной системы предоставления ПЦС на формирование у трудоспособного населения компетенций цифровой экономики с указанием востребованности результатов освоения программы в сфере деятельности соответствующих компаний и готовности к рассмотрению заявок наиболее успешно освоивших образовательную программу граждан на прохождение стажировки и (или) собеседования на предмет трудоустройства путем проставления отметки в профиле программы **(Приложение 2).**

**V.Указание на возможные сценарии профессиональной траектории граждан** по итогам освоения образовательной программы (в соответствии с приложением):

По результатам прохождения образовательной программы возможны следующие сценарии профессиональной траектории развития граждан:

1. Лица, работающие по найму в организации, на предприятии или временно отсутствующие на рабочем месте смогу сохранить текущее место работы, развить свои профессиональные качества, повысить заработную плату и уровень дохода, сменить сферу работы без изменения сферы профессиональной деятельности, а также сохранить и развить свою квалификацию.
2. В рамках освоения данной программы у граждан появится возможность освоить новую востребованную профессию и устроиться на высокооплачиваемую работу, в том числе удалённо, зарегистрироваться в качестве самозанятых (фриланс), зарегистрировать ИП/открыть собственный бизнес.

**VIII.Приложенные Скан-копии утверждённой рабочей программы (Приложение 3).**