

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
ПО ДИСЦИПЛИНЕ**

Математические основы машинного обучения

**Код модуля**  
1160717(1)

**Модуль**  
Основы анализа данных и машинного обучения

**Екатеринбург**

Оценочные материалы составлены автором(ами):

<b>№ п/п</b>	<b>Фамилия, имя, отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Борисов Василий Ильич	кандидат технических наук, без ученого звания	Доцент	радиоэлектроники и телекоммуникаций
2	Созыкин Андрей Владимирович	кандидат технических наук, без ученого звания	Доцент	информационных технологий и систем управления

**Согласовано:**

Управление образовательных программ

Т.Г. Комарова

**Авторы:**

- **Борисов Василий Ильич, Доцент, радиоэлектроники и телекоммуникаций**
- **Созыкин Андрей Владимирович, Доцент, информационных технологий и систем управления**

**1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ Математические основы машинного обучения**

1.	Объем дисциплины в зачетных единицах	9	
2.	Виды аудиторных занятий	Лекции Практические/семинарские занятия	
3.	Промежуточная аттестация	Зачет Экзамен	
4.	Текущая аттестация	Контрольная работа	3
		Домашняя работа	3

**2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Математические основы машинного обучения**

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Планируемые результаты обучения (индикаторы)	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
ОПК-2 -Способен самостоятельно ставить, формализовывать и решать задачи, относящиеся к профессиональной деятельности, используя методы моделирования и математического анализа	Д-1 - Проявлять ответственность и настойчивость в достижении цели З-2 - Характеризовать сферы применения и возможности пакетов прикладных программ для решения задач профессиональной деятельности У-2 - Использовать методы моделирования и математического анализа, в том числе с использованием пакетов прикладных программ для решения задач	Домашняя работа № 1 Домашняя работа № 2 Домашняя работа № 3 Зачет Контрольная работа № 1 Контрольная работа № 2 Контрольная работа № 3 Лекции Практические/семинарские занятия Экзамен

	профессиональной деятельности	
УК-7 -Способен обрабатывать, анализировать, передавать данные и информацию с использованием цифровых средств для эффективного решения поставленных задач с учетом требований информационной безопасности	<p>З-1 - Сделать обзор угроз информационной безопасности, основных принципов организации безопасной работы в информационных системах и в сети интернет</p> <p>З-3 - Сделать обзор современных цифровых средств и технологий, используемых для обработки, анализа и передачи данных при решении поставленных задач</p> <p>П-1 - Обосновать выбор технических и программных средств защиты персональных данных и данных организации при работе с информационными системами на основе анализа потенциальных и реальных угроз безопасности информации</p> <p>У-2 - Выбирать современные цифровые средства и технологии для обработки, анализа и передачи данных с учетом поставленных задач</p>	<p>Домашняя работа № 1</p> <p>Домашняя работа № 2</p> <p>Домашняя работа № 3</p> <p>Зачет</p> <p>Контрольная работа № 1</p> <p>Контрольная работа № 2</p> <p>Контрольная работа № 3</p> <p>Лекции</p> <p>Практические/семинарские занятия</p> <p>Экзамен</p>

### 3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

#### 3.1. Процедуры текущей и промежуточной аттестации по дисциплине

<b>1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5</b>		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>контрольная работа</i>	1,18	100
<b>Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5</b>		
<b>Промежуточная аттестация по лекциям – зачет</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5</b>		

<b>2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5</b>		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	1,18	100
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– <b>1</b>		
Промежуточная аттестация по практическим/семинарским занятиям– <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– <b>не предусмотрено</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –не предусмотрено</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по лабораторным занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – <b>не предусмотрено</b>		
<b>4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено</b>		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по онлайн-занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – <b>не предусмотрено</b>		

### 3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– <b>не предусмотрено</b>		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – <b>не предусмотрено</b>		

#### 3.1. Процедуры текущей и промежуточной аттестации по дисциплине

<b>2. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5</b>
--

Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>контрольная работа</i>	2,18	100
Весовой коэффициент значимости результатов текущей аттестации по лекциям – <b>0.5</b>		
Промежуточная аттестация по лекциям – <b>зачет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – <b>0.5</b>		
<b>2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5</b>		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	2,18	100
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– <b>1</b>		
Промежуточная аттестация по практическим/семинарским занятиям– <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– <b>не предусмотрено</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –не предусмотрено</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по лабораторным занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – <b>не предусмотрено</b>		
<b>4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено</b>		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по онлайн-занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – <b>не предусмотрено</b>		

### 3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах

Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– <b>не предусмотрено</b>
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – <b>не предусмотрено</b>

### 3.1. Процедуры текущей и промежуточной аттестации по дисциплине

<b>3. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5</b>		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>контрольная работа</i>	3,18	100
Весовой коэффициент значимости результатов текущей аттестации по лекциям – <b>0.5</b>		
Промежуточная аттестация по лекциям – <b>экзамен</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – <b>0.5</b>		
<b>2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5</b>		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	3,18	100
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– <b>1</b>		
Промежуточная аттестация по практическим/семинарским занятиям– <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– <b>не предусмотрено</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –не предусмотрено</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по лабораторным занятиям – <b>нет</b>		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – <b>не предусмотрено</b>		
<b>4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено</b>		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям - <b>не предусмотрено</b>		
Промежуточная аттестация по онлайн-занятиям – <b>нет</b>		

**Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено**

### 3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<b>Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено</b>		
<b>Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено</b>		

## 4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

### Критерии оценивания учебных достижений обучающихся

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

### Шкала оценивания достижения результатов обучения (индикаторов) по уровням



<b>Характеристика уровней достижения результатов обучения (индикаторов)</b>				
<b>№ п/п</b>	<b>Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)</b>	<b>Шкала оценивания</b>		
		<b>Традиционная характеристика уровня</b>		<b>Качественная характеристика уровня</b>
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

## **5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ**

### **5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля**

#### **5.1.1. Лекции**

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

#### **5.1.2. Практические/семинарские занятия**

Примерный перечень тем

1. Операции над векторами и матрицами с использованием библиотеки NumPy
2. Работа с Наборами Данных с использованием библиотеки Pandas
3. Разложение матриц с использованием метода главных компонент и сингулярного разложения матрица
4. Предварительная обработка данных
5. Кластеризация данных методом k-средних
6. Реализация алгоритма линейной регрессии
7. Реализация алгоритма логистической регрессии
8. Применение библиотеки sklearn для решения задач регрессии, классификации и кластеризации
9. Кластеризация данных методом DBSCAN

10. Кластеризация данных методом иерархической кластеризации
11. Классификация и Регрессия данных методом опорных векторов
12. Классификация и Регрессия данных методом k-Ближайших соседей
13. Визуализация данных методом Neighborhood Component Analysis
14. Визуализация данных методом t-SNE
15. Классификация данных с использованием Наивного Байесовского классификатора
16. Классификация данных с использованием Дискриминантного Анализа
17. Визуализация данных с использованием Линейного дискриминанта Фишера
18. Классификация и Регрессия данных с использованием Деревьев Решений
19. Классификация и Регрессия данных с использованием ансамблевых методов
20. Продвинутое генерация признаков

Примерные задания

Интеллектуальный анализ данных

Бизнес-аналитика

Перспективные исследования в сфере анализа данных

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

## **5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля**

Разноуровневое (дифференцированное) обучение.

### **Базовый**

#### **5.2.1. Контрольная работа № 1**

Примерный перечень тем

1. История машинного обучения и базовые понятия
2. Данные
3. Линейная Алгебра
4. Методы разложения матриц
5. Предварительная обработка данных
6. Кластеризация
7. Основы математического анализа
8. Регрессия
9. Классификация

Примерные задания

Опишите разницу между подходом машинного обучения и традиционным программированием.

Опишите разницу между задачами классификации и задачами регрессии.

Приведите несколько примеров непрерывных и дискретных данных.

К какому типу данных можно отнести диагноз, поставленный врачом?

У вас есть три матрицы A, B, C: A имеет размеры  $5 \times 4$ , B имеет размеры  $4 \times 6$ , C имеет размеры  $3 \times 5$ . Напишите все возможные матрицы, которые можно перемножить между собой, и укажите размеры результирующих матриц

Что означают собственные значения и собственные векторы ковариационной матрицы в методе главных компонент?

Что означают матрицы  $U$ ,  $S$  и  $V$  в сингулярном разложении матрицы?

Опишите разные ситуации, в которых вы будете использовать разные типы предварительной обработки данных

Какая связь между евклидовым расстоянием и расстоянием Минковского?

Какой гиперпараметр нужно настроить для работы алгоритма кластеризации  $k$ -средних?

Может ли коэффициент силуэта быть равным отрицательному числу? Если «да» - в каких случаях, если «нет» - почему?

В чем заключаются основные различия между методом наименьших квадратов и градиентным спуском для нахождения коэффициентов регрессии?

В каком случае среднеквадратичная логарифмическая ошибка более подходящая метрика, чем среднеквадратичная ошибка?

Может ли коэффициент детерминации быть отрицательным числом? Если «да» - в каких случаях, если «нет» - почему?

Почему L1-регуляризация может привести к отбору значимых признаков (в отличие от L2-регуляризации)?

В чем основное различие между задачами классификации и задачами регрессии?

Допустим, пришли результаты теста  $mr$ . К на коронавирус. Тест дал положительный ответ, хотя на самом деле у  $mr$ . К нет коронавируса. Ошибку какого рода допустил тест?

Как можно построить поверхность принятия решений для логистической регрессии?

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

## 5.2.2. Контрольная работа № 2

Примерный перечень тем

1. Библиотеки Машинного Обучения
2. Продвинутые алгоритмы кластеризации
3. Метод опорных векторов
4. Ближайшие соседи
5. Байесовские методы
6. Деревья Решений
7. Ансамблевые методы
8. Лучшие практики применения методов машинного обучения

Примерные задания

Иерархическая кластеризация: в чем разница между разными типами связей?

DBSCAN: какие точки считаются шумом, граничной точкой, основной точкой?

Какие точки считаются опорными векторами (для задач классификации и регрессии)?

В чем разница между Hard Margin SVM (жестким зазором) и Soft Margin SVM (мягким зазором)?

Почему Kernel trick помогает улучшить результаты метода опорных векторов?

В чем основное отличие использования метода  $k$ -ближайших соседей в при классификации и при регрессии?

Как найти оптимальное значение гиперпараметра  $k$  для методов ближайших соседей?

Как можно уменьшить размерность данных с помощью Nearest Components Analysis?  
Какой гиперпараметр в реализации t-sne связан с балансом между локальными и глобальными аспектами структуры данных?  
Какого рода информацию нужно получить, чтобы использовать теорему Байеса?  
Что означает «наивный» в наивном байесовском классификаторе?  
Предположим есть некое заболевание. Распространенность заболевания 1%. У Вас есть тест, который имеет чувствительность 90% и специфичность 91%. Сколько испытуемых, у которых тест дал положительный результат на самом имеют заболевание?  
В чем разница между линейным и квадратичным дискриминантным анализом?  
Почему линейный дискриминантный анализ может использоваться как метод уменьшения размерности?  
Перечислите основные элементы дерева решений?  
Как выбираются наиболее оптимальные узлы решения?  
В чем разница между использованием деревьев решений при классификации и при регрессии?  
В чем разница между методами бустинга и методами усреднения?  
Что такое «слабый предсказатель» в контексте ансамблевых методов?  
В чем разница в объединении деревьев для алгоритмов Random Forest, Gradient Boosting и AdaBoost?  
В чем заключается основная идея методов фильтрации для выбора значимых параметров?  
Каков основной принцип, лежащий в основе методов Wrapper для выбора значимых параметров?  
Какие концепции необходимы для успешного применения генетического алгоритма для выбора значимых параметров?  
LMS-платформа  
1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

### 5.2.3. Контрольная работа № 3

Примерный перечень тем  
1. Продвинутая генерация признаков  
Примерные задания  
Каким образом категориальные признаки могут быть корректно использованы в линейных моделях?  
Каким образом методы, основанные на деревьях решений, могут использовать комбинаторные признаки?  
LMS-платформа  
1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

### 5.2.4. Домашняя работа № 1

Примерный перечень тем  
1. Основы линейной алгебры в среде Python.  
Примерные задания

а. выполните импорт и визуализацию набора данных с использованием библиотеки Pandas

б. выполните предварительную обработку данных используя векторные и матричные операции с использованием библиотеки NumPy

в. реализуйте алгоритм разложения матриц с использованием библиотеки NumPy

а. выполните реализацию алгоритма кластеризации данных методом k-Средних с использованием библиотеки NumPy

б. выполните реализацию алгоритма линейной регрессии с использованием библиотеки NumPy

в. выполните реализацию алгоритма логистической регрессии с использованием библиотеки NumPy

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

### 5.2.5. Домашняя работа № 2

Примерный перечень тем

1. Алгоритмы кластеризации и визуализации данных в библиотеке sklearn.

2. Алгоритмы регрессии в библиотеке sklearn.

3. Алгоритмы классификации в библиотеке sklearn.

Примерные задания

а. выполните кластеризацию учебного набора данных с использованием метода k-Средних

б. выполните кластеризацию учебного набора данных с использованием метода DBSCAN

в. выполните кластеризацию учебного набора данных с использованием метода иерархической кластеризации

г. выполните визуализацию учебного набора данных методом Neighnorhood Component Analysis

д. выполните визуализацию учебного набора данных методом t-SNE

а. выполните регрессию для учебного набора данных с использованием линейной регрессии

б. выполните регрессию для учебного набора данных с использованием метода ближайших соседей

в. выполните регрессию для учебного набора данных с использованием метода опорных векторов

г. выполните регрессию для учебного набора данных с использованием деревьев решений

д. выполните регрессию для учебного набора данных с использованием ансамблевых методов

а. выполните классификацию для учебного набора данных с использованием логистической регрессии

б. выполните классификацию для учебного набора данных с использованием метода ближайших соседей

в. выполните классификацию для учебного набора данных с использованием метода опорных векторов

г. выполните классификацию для учебного набора данных с использованием деревьев решений

д. выполните классификацию для учебного набора данных с использованием ансамблевых методов

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

### 5.2.6. Домашняя работа № 3

Примерный перечень тем

1. Методы продвинутой генерации признаков.

2. Применение алгоритмов Машинного обучения для решения задач уменьшения размерности, кластеризации, регрессии и классификации.

Примерные задания

а. выполните генерацию дополнительных категориальных признаков с использованием onehot encoding для реального набора данных

б. выполните генерацию дополнительных числовых признаков с использованием mean encoding для реального набора данных

а. Приведите описание набора данных. Описание должно как минимум включать название набора данных, краткую аннотацию, количество параметров и экземпляров в наборе данных. Желательно добавлять небольшие таблицы с примерами

б. Для представления результаты алгоритмов уменьшения размерности необходимо включить представление исходного набора данных в новом пространстве, выделив одни и те же классы отдельными маркерами. Также рекомендуется представить матрицы преобразования, чтобы увидеть какие параметры каждый метод использует. Для t-sne представьте несколько результатов для разных значений perplexity

с. Представьте результаты наиболее оптимальной кластеризации, указав гиперпараметры, которые использовались для получения этого оптимального разделения

д. Представьте результаты оптимальной классификации, указав гиперпараметры, которые использовались для получения этой оптимальной классификации. Представьте метрики классификации как для обучающих, так и для тестовых данных. Выполните визуализацию в зависимости от используемого алгоритма.

е. Представьте результаты оптимальной регрессии, указав гиперпараметры, которые использовались для получения этой оптимальной регрессии. Представьте метрики регрессии как для обучающих, так и для тестовых данных. Выполните визуализацию в зависимости от используемого алгоритма.

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

### 5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

#### 5.3.1. Зачет

Список примерных вопросов

1. Типы задач машинного обучения
  2. Типы данных
  3. Недостаточное обучение и переобучение (Underfitting и Overfitting)
  4. Градиентный спуск
  5. Перекрестная проверка
  6. Матрица ошибок и метрики классификации
  7. Метрики регрессии
  8. Предварительная обработка данных
  9. Кластеризация k-средних
  10. Иерархическая кластеризация
  11. DBSCAN
  12. Метрики кластеризации
  13. Метод главных компонент
  14. Сингулярное Разложение Матриц
  15. Neighborhood Components Analysis
  16. Визуализация данных методом t-SNE
  17. Применение Линейного дискриминантного анализа для уменьшения размерности
  18. Линейная регрессия
  19. Регуляризация линейной регрессии
  20. Регрессия k-ближайших соседей
  21. Регрессия деревьев решений
  22. Метод опорных векторов для регрессии
  23. Регрессия с использованием AdaBoost
  24. Регрессия с использованием Gradient Boosting
  25. Регрессия с использованием Random Forest
  26. Логистическая регрессия
  27. Классификатор k-ближайших соседей
  28. Наивный байесовский классификатор
  29. Дискриминантный анализ (интерпретация Фишера)
  30. Дискриминантный анализ (байесовская версия)
  31. Классификация с использованием деревьев решений
  32. Метод опорных векторов (soft и hard margin)
  33. Метод опорных векторов (kernel trick)
  34. Классификация с использованием AdaBoost
  35. Классификация с использованием Gradient Boosting
  36. Классификация с использованием Random Forest
- LMS-платформа – не предусмотрена

#### 5.3.2. Экзамен

Список примерных вопросов

1. Ключевые шаги (pipeline) применения машинного обучения для анализа данных. Описание конкретного примера.
  2. Ошибки моделей. Компромисс смещения и дисперсии. Ансамблевые модели машинного обучения.
  3. Переобучение и недообучение. Методы перекрестной проверки. Особенности разбиения выборки для задач регрессии, классификации и временных рядов.
  4. Градиентный спуск. Применение градиентного спуска в линейной регрессии, Neighborhood Components Analysis, t-SNE.
  5. Интерпретация моделей машинного обучения. Описание примеров для различных подходов.
  6. Параметры и гиперпараметры моделей машинного обучения. Подходы к оптимизации гиперпараметров моделей.
  7. Генерация новых параметров на основе моделей машинного обучения.
  8. Предварительная обработка данных.
  9. Генерация дополнительных параметров.
  10. Обучение модели.
  11. Проверка достоверности результатов модели.
  12. Анализ полученной модели.
  13. Ключевые шаги (pipeline) применения машинного обучения для анализа данных. Описание конкретного примера.
- LMS-платформа – не предусмотрена

#### **5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности**

Направления воспитательной деятельности сопрягаются со всеми результатами обучения компетенций по образовательной программе, их освоение обеспечивается содержанием всех дисциплин модулей.