

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Машинное обучение

Код модуля
1160955(1)

Модуль
Машинное обучение

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Долганов Антон Юрьевич	кандидат технических наук, без ученого звания	Доцент	радиоэлектроники и телекоммуникаций

Согласовано:

Управление образовательных программ

Т.Г. Комарова

Авторы:

- Долганов Антон Юрьевич, Доцент, радиоэлектроники и телекоммуникаций

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ Машинное обучение

1.	Объем дисциплины в зачетных единицах	6	
2.	Виды аудиторных занятий	Лекции Лабораторные занятия	
3.	Промежуточная аттестация	Зачет Экзамен	
4.	Текущая аттестация	Контрольная работа	2
		Домашняя работа	7

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Машинное обучение

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 П-1 - Решать самостоятельно сформулированные практические задачи, относящиеся к профессиональной деятельности методами моделирования и математического анализа, в том числе с использованием пакетов прикладных программ	Домашняя работа № 1 Домашняя работа № 2 Домашняя работа № 3 Домашняя работа № 4 Домашняя работа № 5 Домашняя работа № 6 Домашняя работа № 7 Зачет Контрольная работа № 1 Контрольная работа № 2 Лабораторные занятия Лекции Экзамен

	У-2 - Использовать методы моделирования и математического анализа, в том числе с использованием пакетов прикладных программ для решения задач профессиональной деятельности	
ПК-1 -Способен использовать методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации посредством информационных технологий	<p>З-2 - Характеризовать особенности применения информационных технологий в области получения, передачи, хранения, переработки и представления информации</p> <p>П-1 - Осуществлять обоснованный выбор методов цифровой обработки сигналов для переработки и представления информации посредством информационных технологий с учетом особенностей сигнала</p> <p>У-1 - Выбирать адекватные методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации</p>	<p>Домашняя работа № 1</p> <p>Домашняя работа № 2</p> <p>Домашняя работа № 3</p> <p>Домашняя работа № 4</p> <p>Домашняя работа № 5</p> <p>Домашняя работа № 6</p> <p>Домашняя работа № 7</p> <p>Зачет</p> <p>Контрольная работа № 1</p> <p>Контрольная работа № 2</p> <p>Лабораторные занятия</p> <p>Лекции</p> <p>Экзамен</p>
ПК-2 -Способен разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений	<p>З-1 - Интерпретировать результаты применения математических моделей процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений</p> <p>П-1 - Иметь практический опыт применения математического моделирования и анализа данных</p> <p>У-1 - Определять оптимальные математические модели для решения задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений в зависимости от заданных условий</p>	<p>Домашняя работа № 1</p> <p>Домашняя работа № 2</p> <p>Домашняя работа № 3</p> <p>Домашняя работа № 4</p> <p>Домашняя работа № 5</p> <p>Домашняя работа № 6</p> <p>Домашняя работа № 7</p> <p>Зачет</p> <p>Контрольная работа № 1</p> <p>Контрольная работа № 2</p> <p>Лабораторные занятия</p> <p>Лекции</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

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.6		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>контрольная работа №1</i>	2,16	100
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – зачет		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – не предусмотрено		
Промежуточная аттестация по практическим/семинарским занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – не предусмотрено		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – 0.4		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа №1</i>	2,16	25
<i>домашняя работа №2</i>	2,16	25
<i>домашняя работа №3</i>	2,16	25
<i>домашняя работа №4</i>	2,16	25
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям – 1		
Промежуточная аттестация по лабораторным занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено		
4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий – не предусмотрено		
Текущая аттестация на онлайн-занятиях	Сроки – семестр,	Максимальная оценка в баллах

	учебная неделя	
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено		
Промежуточная аттестация по онлайн-занятиям –нет		
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
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

2. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.6		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>контрольная работа №2</i>	3,16	100
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – экзамен		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– не предусмотрено		
Промежуточная аттестация по практическим/семинарским занятиям–нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– не предусмотрено		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –0.4		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа №5</i>	3,16	30
<i>домашняя работа №6</i>	3,16	35
<i>домашняя работа №7</i>	3,16	35

Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям -1		
Промежуточная аттестация по лабораторным занятиям –нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено		
4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено		
Промежуточная аттестация по онлайн-занятиям –нет		
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
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

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	<p>Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов.</p> <p>Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения.</p> <p>Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.</p>
-------------------	---

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

Характеристика уровней достижения результатов обучения (индикаторов)				
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
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. Лабораторные занятия

Примерный перечень тем

1. Предварительная обработка данных. Библиотека Pandas
2. Алгоритмы кластеризации и уменьшения размерности. Библиотека scikit-learn
3. Алгоритмы регрессии. Библиотека scikit-learn
4. Алгоритмы классификации. Библиотека scikit-learn
5. Методы компьютерного зрения. Библиотека OpenCV
6. Нейронные сети. Библиотека TensorFlow
7. Обработка естественного языка. Библиотека HuggingFace

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

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. Базовые понятия машинного обучения
2. Типы данных
3. Классические алгоритмы машинного обучения

Примерные задания

Вы выполнили One-hot кодирование категориальной переменной со следующими возможными значениями {red,blue,pink,yellow, green}. Сколько новых столбцов данных вам для этого потребовалось?

Укажите правильные утверждения о предварительной обработке данных

После нормализации среднее значение параметра как правило выше, чем после стандартизации

После нормализации среднее значение параметра как правило ниже, чем после стандартизации

После нормализации максимальное значение параметра как правило выше, чем после стандартизации

После нормализации максимальное значение параметра как правило ниже, чем после стандартизации

Есть тест на некое заболевание RjDbL-2077. Мы хотим отслеживать заболевших. Тест дал положительный ответ, хотя на самом деле у испытуемого нет заболевания. Какую ошибку допустил тест (перечислите все возможные варианты ответов)

Ошибка первого рода

Ошибка второго рода

Пропуск

Ложное срабатывание

Допустим, есть два классификатора: первый классификатор имеет точность 95%, чувствительность 99%, специфичность 50%; второй классификатор имеет точность 87%,

чувствительность 84%, специфичность 94%. Что вы можете сказать о данных, используемых для классификации? Какой из этих классификаторов надежнее (при условии, что нам важно определение обоих классов)?

Вероятнее всего данные сбалансированы по классам

Вероятнее всего данные не сбалансированы по классам

Первый классификатор надежнее

Второй классификатор надежнее

Выберите возможные гиперпараметры модели логистической регрессии

w веса признаков

λ константа регуляризации

тип регуляризации

порог значений сигмоиды

b_0 смещение

Вы решаете задачу классификации с использованием логистической регрессии.

Целевая переменная состоит из 4 классов. Для какой стратегии вам потребуется использовать меньше бинарных классификаторов

Один против всех

Один против одного

Может ли коэффициент детерминации быть отрицательным числом

Есть три центроида: $C1 \{1,0,0\}$, $C2 \{0,1,1\}$, $C3 \{1,0,1\}$. Есть точка x с координатами $(2,0,2)$. К какому кластеру следует отнести эту точку, при использовании Евклидовой метрики расстояния?

$C1$

$C2$

$C3$

Как называется метод определения оптимального числа k (кластеров) с использованием анализа инерции?

- Метод Пятки
- Метод Плеча
- Метод Колена
- Метод Локтя

Какие значения k в методе ближайших соседей более устойчивы к шумовым данным?

- Небольшие значения k
- Большие значения k

Представьте, что вы обучили модель Случайный Лес, в котором 50 деревьев.

Рассмотрим 2 случая:

1) Мы убираем первое дерево из модели

2) Мы убираем последнее дерево из модели.

Предположите, какое утверждение справедливо при оценке качества моделей на тренировочных данных

Качество моделей будет примерно одинаковым

Качество моделей в первом случае будет хуже

Качество моделей во втором случае будет хуже

Представьте, что вы обучили модель с использованием Градиентного Бустинга, в котором 50 деревьев. Рассмотрим 2 случая:

- 1) Мы убираем первое дерево из модели
- 2) Мы убираем последнее дерево из модели.

Предположите, какое утверждение справедливо при оценке качества моделей на тренировочных данных

- Качество моделей будет примерно одинаковым
- Качество моделей в первом случае будет хуже
- Качество моделей во втором случае будет хуже

Рассмотри две ситуации.

1) Вы обучили две модели классификации Случайный Лес, по 25 деревьев в каждой. Ваше итоговое предсказание – усреднение двух моделей

- 2) Вы обучили одну модель классификации Случайный Лес, в которой 50 деревьев. Остальные гиперпараметры моделей совпадают

Предположите, какое утверждение справедливо при оценке качества моделей на тренировочных данных

- Качество моделей будет примерно одинаковым
- Качество моделей в первом случае будет хуже
- Качество моделей во втором случае будет хуже

Какие утверждения справедливы для Моделей Случайный лес и Градиентный Бустинг
Деревья в Случайном лесе могут обучаться параллельно
Деревья в Случайном лесе обучаются последовательно
Деревья в Градиентный Бустинге могут обучаться параллельно
Деревья в Градиентный Бустинге обучаются последовательно

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.2. Контрольная работа № 2

Примерный перечень тем

1. Компьютерное зрение
2. Нейронные сети
3. Обработка естественного языка

Примерные задания

В чем принципиальная разница между цветовыми моделями RGB и CMYK?

В чем преимущество вспомогательных моделей (HSI, HSV) по сравнению с RGB?

Какая идеология стоит за методом настройки цветового баланса «Серый мир»?

Какой метод бинаризации следует использовать, чтобы получить исходный объект на черном фоне?

Почему рекомендуется фильтровать изображение перед поиском границ?

Для каких задач поиск особых точек на изображении может быть полезен?

Почему метод SURF работает быстрее SIFT?

В чем основное отличие нейросетевого подхода от «классического» машинного обучения?

Каковы ключевые моменты алгоритма обратного распространения ошибки при обучении нейронной сети?

В чем отличие Sequential API и Functional API в TensorFlow?

Какие ключевые методы используются для предотвращения переобучения нейронных сетей?

Что является обучаемыми параметрами (весами) при использовании сверточных нейронных сетей?

Какие задачи лучше всего подходят для использования сверточных слоев, а какие - для полносвязных слоев?

В чем основные различия между базой данных ImageNet и базой данных CIFAR-10?

Что делает свертка 1×1 и что показывают веса обученных сверток 1×1 ?

Какая особенность архитектуры ResNet позволила преодолеть проблему исчезающего градиента?

Каким образом в архитектуре MobileNet уменьшается количество весов и математических операций?

Какие нововведения были использованы в YOLOv2 по сравнению с оригинальной версией?

Каков основной принцип работы генеративных состязательных сетей (GAN)?

Каковы основные блоки Авто-Энкодеров?

Почему увеличение N в моделях N -грамм помогает улучшить качество сгенерированного текста? Как это можно измерить?

В чем заключается идея использования Stemming в качестве метода предварительной обработки текста?

В чем заключаются основные идеи модели векторного пространства Word2Vec?

Создайте разные нейронные сети с помощью библиотеки tensorflow

Структура нейронной сети: слой Embedding и либо 1D свёртка, либо один из вариантов рекуррентного слоя (SimpleRNN, GRU, LSTM)

Укажите число обучаемых параметров если:

размер словаря 5000

размерность embedding 8

размерности рекуррентных слоев 2

Количество 1D фильтров 5, размер свертки 3

Какого цвета кукла из «Улицы Сезам», в честь которой назван подход построения признакового пространства (эмбеддинга) текстов на основе двунаправленных рекуррентных нейронных сетей.

Почему есть причина находить баланс между токенами, подобными алфавиту (каждый отдельный символ является токеном), и токенами, подобными словарю (каждое отдельное слово является токеном)?

Что такое Key, Values and Query когда мы говорим о механизме внимания?

Опишите основные различия между моделями, подобными BERT, и моделями, подобными GPT?

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.3. Домашняя работа № 1

Примерный перечень тем

1. Обработка и визуализация данных в библиотеке Pandas

2. Основы работы с Наборами Данных

3. Предварительная обработка Данных

Примерные задания

Основы работы с Наборами Данных

Найдите и загрузите несколько интересных наборов данных из OpenML (или любой другой сайт с данными. Можете "что-то свое"). Это пригодится на будущее — вам понадобится набор данных для регрессии, классификации, кластеризации и уменьшения размерности.

Упаковать набор данных в ДатаФрейм pandas с Именованными столбцами

Выполните Расчет статистик (в т.ч. с использованием агрегации)

Нарисуйте как минимум 3 разные графика (по крайней мере, на одном графике вам нужно сделать цвет или размер маркеров на основе целевого класса / значений)

Предварительная обработка Данных

Для выбранных наборов данных выберите признаки, которые необходимо предварительно обработать.

Выполните адекватную предварительную обработку данных

Проанализируйте результат

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.4. Домашняя работа № 2

Примерный перечень тем

1. Алгоритмы классификации. Библиотека scikit-learn

2. Логистическая Регрессия

3. Ближайшие Соседи

4. Байесовские Методы

5. Метод опорных векторов

6. Деревья Решений

7. Ансамблевые методы

Примерные задания

Логистическая Регрессия

• Выберите набор данных классификации из OpenML для анализа (предпочтительна бинарная классификация)

* Вы можете использовать методы уменьшения размерности

• Выполните классификацию с разными подходами

* Вы можете использовать оригинальные параметры

* Вы можете использовать полиномиальные параметры

* Вы можете использовать параметры после применения уменьшения размерности

• Оцените показатели классификации с помощью перекрестной проверки и матрицы ошибок

• Визуализируйте результаты

Классификация k-Ближайших Соседей

• Выберите набор данных классификации из OpenML для анализа

• Выполните классификацию k-Ближайших соседей. Выберите оптимальные гиперпараметры

• Сравните результаты с логистической регрессией

Наивный Байесовский классификатор

- Выберите набор данных классификации из OpenML для анализа
- Выполните классификацию Наивным Байесовским классификатором. Выберите оптимальные гиперпараметры (априорные вероятности)
- Сравните результаты с логистической регрессией (и другими классификаторами)

Дискриминантный Анализ (Классификация)

- Выберите набор данных регрессии из OpenML для анализа
- Выполните классификацию Дискриминантным Анализом. Выберите оптимальные гиперпараметры (альфы там, линейный или квадратичный)
- Сравните результаты с логистической регрессией (и другими классификаторами)

Методом Опорных Векторов

- Выберите набор данных классификации из OpenML для анализа
- Выполните классификацию Методом Опорных Векторов. Выберите оптимальные гиперпараметры
- Сравните результаты с логистической регрессией (и прочими методами)

Классификация Деревьями Решений

- Выберите набор данных классификации из OpenML для анализа
- Выполните классификацию Деревьями Решений. Выберите оптимальные гиперпараметры
- Сравните результаты с логистической регрессией (и прочими методами)

Классификация Ансамблями

- Выберите набор данных классификации из OpenML для анализа
- Выполните классификацию Ансамблями (выбирете как минимум два вида). Выберите оптимальные гиперпараметры
- Сравните результаты с логистической регрессией (и прочими методами)

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.5. Домашняя работа № 3

Примерный перечень тем

1. Алгоритмы кластеризации и уменьшения размерности. Библиотека scikit-learn
2. Кластеризация K-Средних
3. Иерархическая Кластеризация
4. Кластеризация DBSCAN
5. Метод главных компонент (PCA)
6. Neighbourhood Components Analysis
7. tSNE

Примерные задания

Кластеризация

- Выберите набор данных для кластеризации или классификации из OpenML для анализа (это может быть 2-х или 3-х мерный набор данных, или вы можете использовать только 2/3 оси для визуализации)

- * Рекомендуется выполнить стандартизацию данных до применения кластеризации.

- * Вы можете использовать уменьшение размерности (PCA) вместо исходных параметров

- Выполните кластеризацию с использованием различных подходов

- * Настройте гиперпараметры различных подходов к кластеризации для получения лучших результатов

- Визуализируйте результаты

PCA

- Выберите набор данных с сайта OpenML (он должен иметь > 10 параметров, как минимум 2 класса и не слишком много образцов (менее 10000))

- Примените PCA

- Визуализируйте несколько различных главных компонент (вы можете использовать двухмерные или трехмерные графики и различные комбинации главных компонент, такие как pca-1 pca-2 pca-5; pca-2 pca-3 pca-4; pca-1 pca-5 pca-9)

- Визуализируйте веса, чтобы понять, что означают различные основные компоненты.

Сделайте краткий анализ

Neighbourhood Components Analysis

- Выберите набор данных с сайта OpenML (он должен иметь > 10 параметров, как минимум 2 класса и не слишком много образцов (менее 10000))

- Выполните Neighbourhood Components Analysis

- Визуализируйте вложения и веса вложений. Сделайте краткий анализ

- Сравните с результатами PCA / SVD

tSNE

- Выберите набор данных с сайта OpenML (он должен иметь > 10 параметров, как минимум 2 класса и не слишком много образцов (менее 10000))

- Выполните tSNE несколько раз (поскольку оно случайное по своей природе)

- Визуализируйте вложения. Сделайте краткий анализ (насколько стабильны результаты, например)

- Сравните с результатами PCA / SVD

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.6. Домашняя работа № 4

Примерный перечень тем

1. Алгоритмы регрессии. Библиотека scikit-learn

2. Линейная Регрессия

3. Ближайшие Соседи

4. Метод опорных векторов

5. Деревья Решений

6. Ансамблевые методы

Примерные задания

Линейная Регрессия

- Выберите набор данных регрессии из OpenML для анализа
- Выполните регрессию с помощью разных подходов:
 - * Вы можете использовать простую линейную модель
 - * Вы можете использовать только регуляризацию
 - * Вы можете комбинировать регуляризацию и полиномиальные параметры
- Оцените метрики регрессии с помощью перекрестной проверки
- Визуализируйте результаты (веса, предсказания, и т.п.)

Регрессия к-Ближайших Соседей

- Выберите набор данных регрессии из OpenML для анализа
- Примените Регрессия к-Ближайших Соседей. Выберите оптимальные гиперпараметры
- Сравните результаты с линейной регрессией

Регрессия Методом Опорных Векторов

- Выберите набор данных регрессии из OpenML для анализа
- Примените Регрессия Методом Опорных Векторов. Выберите оптимальные гиперпараметры
- Сравните результаты с линейной регрессией (и прочими методами)

Регрессия Деревьями Решений

- Выберите набор данных регрессии из OpenML для анализа
- Примените Регрессия Деревьями Решений. Выберите оптимальные гиперпараметры
- Сравните результаты с линейной регрессией (и прочими методами)

Регрессия Ансамблями

- Выберите набор данных регрессии из OpenML для анализа
- Примените Регрессию Ансамблями (выберете как минимум два вида). Выберите оптимальные гиперпараметры
- Сравните результаты с линейной регрессией (и прочими методами)

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.7. Домашняя работа № 5

Примерный перечень тем

1. Методы компьютерного зрения. Библиотека OpenCV
2. Компьютерное Зрение: Базовые Операции
3. Компьютерное Зрение: Продвинутое Операции
4. Компьютерное Зрение: Особые Точки

Примерные задания

Коррекция яркости и баланс цветов

- Сделать фото в не очень хороших условиях (чтобы были явные проблемы с освещением).

- Применить операции для корректировки изображения.

Бинаризация

- Выбрать простое изображение по типу "красные помидору на фоне зеленых".
- Попытайтесь сделать бинаризацию "руками" чтобы отделить одни объекты от других (целевой объект и фон)
- Реализовать бинаризацию по методу Отсу для каждого канала отдельно (или для определенных каналов).
- Возможно предстоит сменить цветовую модель для наилучшего эффекта

Сегментация

- Выберите изображение на котором несколько однотипных объектов (монеты, карты, камни) и т.д.
- Выполните его сегментацию

Мешок Слов

- Возьмите несколько разных изображений (не менее 3) в хорошем качестве.
- Например: какое-нибудь животное, какой-нибудь инструмент и еще что-нибудь
- Выполните построение визуальных слов

Каскады

- Выберите каскад
- Выберите изображение, где могут быть несколько целевых элементов
- Применить каскад и визуализировать результат

Особые точки

- Выберите два изображения - одно изображение на котором вы будете искать ключевые точки, второе - измененное первое (или первое встречается на втором как часть).
- Например: изображение обложки книги и фотография, где встречается эта книга.
- Сравните разные дескрипторы/детекторы (подобрав оптимальные параметры) и сопоставьте изображения

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.8. Домашняя работа № 6

Примерный перечень тем

1. Нейронные сети. Библиотека TensorFlow
2. Полносвязные и сверточные нейронные сети
3. Генеративные нейронные сети

Примерные задания

Классификация Изображений

- Выбрать набор данных изображений из коллекции TensorFlow
- Соберите Свою Нейронную сеть

- Попробуйте классифицировать изображения чтобы было "неплохо" (хотя бы accuracy ~ 70%)

- Визуализируйте Веса Сверточных сетей

Перенос обучения

- Выбрать набор данных изображений из коллекции TensorFlow
- Выберите предварительно обученную модель
- Обучите последний слой, чтобы классифицировать новый набор данных / выполните

Fine-Tune модели для достижения хороших результатов

- Представьте результаты предсказания на новых изображениях

Перенос Стиля

- Выбрать изображения (можно условно даже самим что-то нарисовать) и выбрать стили (условно свой любимый художник)

- Применить Перенос стиля

- Сделать не менее 3 красивых картинок

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

5.2.9. Домашняя работа № 7

Примерный перечень тем

1. Обработка естественного языка. Библиотека HuggingFace

2. Векторные модели

3. Рекуррентные нейронные сети

4. Трансформеры

Примерные задания

Word2Vec

- Придумать 10 различных интересных стилистических, синтаксических, числовых, семантических, географических и т.д. связей векторных представлений

- Визуализировать несколько примеров в пространстве Главных компонент

Генерация Текста с помощью RNN

- Собрать свой текст (можно условного Льва Николаевича Толстого, можно подборку песен вашего любимого исполнителя. Главное, чтобы было много букв)

- Поварьюйте настройки модели (длина последовательности, размерности слоя Embedding, сколько чего в RNN слое, тип слоя)

- Обучить уверенную модель (увеличив количество эпох)

- Посмотреть, что будет при разных начальных строках

- Сравнить для разных значений температуры

Сравнение моделей и Задач на платформе Hugging Face

- Выберите несколько NLP задач (не меньше 4) и выберите разные модели (по 3 для задачи)

- Сравните на одинаковых входных данных результат работы

LMS-платформа

1. <https://elearn.urfu.ru/course/view.php?id=6550>

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. Градиентный спуск
 4. Перекрестная проверка. Оптимизация гиперпараметров моделей
 5. Предварительная обработка данных
 6. Инженерия Признаков
 7. Pipeline в Машинном Обучении
 8. Кластеризация k-средних
 9. Иерархическая кластеризация
 10. DBSCAN
 11. Метод главных компонент
 12. Сингулярное Разложение Матриц
 13. t-SNE
 14. Neighborhood Components Analysis
 15. Линейная регрессия
 16. Регуляризация линейной регрессии
 17. Метод k-ближайших соседей
 18. Метод опорных векторов
 19. Деревья решений
 20. AdaBoost
 21. Gradient Boosting
 22. Логистическая регрессия
 23. Random Forest
 24. Наивный байесовский классификатор
 25. Дискриминантный анализ
- LMS-платформа – не предусмотрена

5.3.2. Экзамен

Список примерных вопросов

1. Базовые понятия в Машинном обучении
2. Классические алгоритмы Машинного обучения
3. Алгоритм Виолы-Джонса для распознавания лиц
4. Обнаружение краев на изображении
5. Цветовые модели
6. Детектор и дескриптор SIFT
7. BRIEF детектор и дескриптор
8. Детектор и дескриптор SURF
9. Визуальный словарь и метод мешка слов

10. Основные элементы полносвязных нейронных сетей: нейроны, функции активации, веса и смещения
 11. Основные элементы сверточных нейронных сетей: свертки, padding, stride, pooling
 12. Оптимизаторы нейронных сетей: градиентный спуск, стохастический градиентный спуск, экспоненциально взвешенные средние, RMSProp, Adam
 13. Особенности нейросетевых архитектур для классификации изображений (Общая структура, AlexNet, VGG)
 14. свертка 1x1 и архитектура нейронной сети Google Inception для классификации изображений
 15. Остаточные блоки и архитектура нейронной сети ResNet для классификации изображений
 16. Depthwise Separable Convolution и архитектура нейронной сети MobileNet для классификации изображений
 17. Развитие архитектур YOLO для обнаружения объектов
 18. Генеративно-сопоставительные сети
 19. Автоэнкодеры и вариационные автоэнкодеры
 20. Передача стиля с помощью нейронных сетей
 21. Предварительная обработка текста
 22. Анализ тональности текстов с помощью машинного обучения
 23. Модель векторного пространства
 24. Наивный машинный перевод
 25. Автозамена и edit distance
 26. Автозаполнение и модели языка N-Gram
 27. Рекуррентные нейронные сети
 28. Gated Recurrent Units и Long Short-Term Memory
 29. Трансформаторы: внимание, multi-head attention, энкодеры и декодеры
 30. Transfer Learning в NLP и языковых моделях
- 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.