

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ,
МИНИСТЕРСТВО НАУКИ, ВЫСШЕГО ОБРАЗОВАНИЯ И ИННОВАЦИЙ
КЫРГЫЗСКОЙ РЕСПУБЛИКИ**

ГОУ ВПО Кыргызско-Российский Славянский университет
имени первого Президента Российской Федерации Б.Н. Ельцина



Машинное обучение и анализ данных рабочая программа дисциплины (модуля)

Закреплена за кафедрой **Информационных и вычислительных технологий**
Учебный план g09040440_24_0 пи_рпис.plx
Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия
Магистерская программа "Разработка программно-информационных систем"

Квалификация **магистр**
Форма обучения **очная**
Общая трудоемкость **4 ЗЕТ**
Часов по учебному плану 144
Виды контроля в семестрах: в том числе: зачеты с оценкой 3
аудиторные занятия 54
самостоятельная работа 89,8

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	3 (2.1)		Итого	
	уп	рп	уп	рп
Неделя	17			
Вид занятий	уп	рп	уп	рп
Лекции	18	18	18	18
Практические	36	36	36	36
Контактная работа в период теоретического обучения	0,2	0,2	0,2	0,2
В том числе инт.	8	8	8	8
В том числе в форме практ.подготовки	36	36	36	36
Итого ауд.	54	54	54	54
Контактная работа	54,2	54,2	54,2	54,2
Сам. работа	89,8	89,8	89,8	89,8
Итого	144	144	144	144

Программу составил(и):

к.т.н., доцент, Верзунов С.Н.; д.т.н., профессор, Лыченко Н.М.



Рабочая программа дисциплины

Машинное обучение и анализ данных

разработана в соответствии с ФГОС 3++:

Федеральный государственный образовательный стандарт высшего образования - магистратура по направлению подготовки 09.04.04 Программная инженерия (приказ Минобрнауки России от 19.09.2017 г. № 932)

составлена на основании учебного плана:

Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия Магистерская программа "Разработка программно-информационных систем"

утвержденного учёным советом вуза от 22.10.2024 протокол № 2.

Рабочая программа одобрена на заседании кафедры

Информационных и вычислительных технологий

Протокол от 02.10.2024 г. № 2

Срок действия программы: 2024-2028 уч.г.

Зав. кафедрой д.т.н., проф. Лыченко Н.М.



Визирование РПД для исполнения в очередном учебном году

Председатель УМС

9 сентября 2025 г.



Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2025-2026 учебном году на заседании кафедры

Информационных и вычислительных технологий

Протокол от 3 сентября 2025 г. № 1
Зав. кафедрой д.т.н., проф. Лыченко Н.М.



Визирование РПД для исполнения в очередном учебном году

Председатель УМС

_____ 2026 г.

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2026-2027 учебном году на заседании кафедры

Информационных и вычислительных технологий

Протокол от _____ 2026 г. № ____
Зав. кафедрой д.т.н., проф. Лыченко Н.М.

Визирование РПД для исполнения в очередном учебном году

Председатель УМС

_____ 2027 г.

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2027-2028 учебном году на заседании кафедры

Информационных и вычислительных технологий

Протокол от _____ 2027 г. № ____
Зав. кафедрой д.т.н., проф. Лыченко Н.М.

Визирование РПД для исполнения в очередном учебном году

Председатель УМС

_____ 2028 г.

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2028-2029 учебном году на заседании кафедры

Информационных и вычислительных технологий

Протокол от _____ 2028 г. № ____
Зав. кафедрой д.т.н., проф. Лыченко Н.М.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	
1.1	формирование у студентов представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining),
1.2	изучение основных подходов и алгоритмов решения задач анализа данных и особенностей их применения к решению реальных задач,
1.3	получение студентами навыка по выявлению, формализации и успешному решению практических задач анализа данных, возникающие в процессе их профессиональной деятельности.
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП	
Цикл (раздел) ООП:	Б1.В
2.1	Требования к предварительной подготовке обучающегося:
2.1.1	Научно-исследовательский практикум
2.1.2	Научно-исследовательская работа
2.1.3	Имитационное моделирование
2.1.4	Системы искусственного интеллекта
2.1.5	Цифровая обработка сигналов
2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:
2.2.1	Выполнение и защита выпускной квалификационной работы
2.2.2	Подготовка к сдаче и сдача государственного экзамена
2.2.3	Преддипломная практика
3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
ПК-4: Владение навыками создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов	
Знать:	
Уровень 1	методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов.
Уровень 2	Иметь представления об экспертных системах, теории распознавания образов, методах и средствах интеллектуального анализа данных
Уметь:	
Уровень 1	использовать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов.
Уровень 2	Применять на практике инструментальные средства систем искусственного интеллекта, строить формальное описание предметной области
Владеть:	
Уровень 1	методами и средствами создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов с помощью современных компьютерных технологий.
Уровень 2	Инструментарием в программировании экспертных систем и реализации методов интеллектуального анализа данных
В результате освоения дисциплины обучающийся должен	
3.1	Знать:
3.1.1	принципы обработки массивов данных, способы их представления и хранения;
3.1.2	основные задачи и методы интеллектуального анализа данных;
3.1.3	математические модели, лежащие в основе различных подходов к решению задач анализа данных
3.1.4	возможности современных и перспективных средств разработки программных продуктов, технических средств;
3.1.5	структуру и содержимое программных пакетов для анализа данных (например sklearn в python).
3.2	Уметь:
3.2.1	формулировать задачи анализа данных;
3.2.2	выбирать адекватные алгоритмы их решения;
3.2.3	выполнять процедуры проектирования хранилищ данных и заполнения готовых хранилищ данными;
3.2.4	оценивать качество получаемых решений;
3.2.5	выбирать средства реализации требований к программному обеспечению.
3.3	Владеть:

3.3.1	технологиями разработки алгоритмов и программными системами анализа данных;							
3.3.2	средствами автоматизации интеллектуального анализа и обработки данных;							
3.3.3	формирование и предоставление отчетности в соответствии с установленными регламентами.							
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)								
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература	Инте ракт.	Пр. подг.	Примечание
	Раздел 1. Введение в машинное обучение							
1.1	Введение в машинное обучение. Основные понятия /Лек/	3	2	ПК-4	Л1.1 Л1.2Л2.1Л3.1 Л3.2 Э2 Э3 Э4 Э5			
1.2	Практическая работа №1. Обучение классификатора цифр /Пр/	3	9	ПК-4	Л1.1 Л1.3Л2.1Л3.2 Э2 Э3 Э4 Э5	2	9	Интерактивный разбор примеров
1.3	Признаки и их виды /Ср/	3	4	ПК-4	Л1.1 Л1.2Л2.1Л3.2 Э2 Э3 Э4 Э5			
1.4	Принцип максимума правдоподобия и его связь с минимизацией эмпирического риска /Ср/	3	4	ПК-4	Л1.1 Л1.2Л2.1Л3.2 Э2 Э3 Э4 Э5			
1.5	Проблема переобучения и обобщающая способность алгоритма /Ср/	3	4	ПК-4	Л1.1 Л1.2Л2.1Л3.2 Э2 Э3 Э4 Э5			
1.6	Выбор алгоритма в вероятностной постановке задачи. /Ср/	3	4	ПК-4	Л1.1 Л1.2Л2.1Л3.2 Э2 Э3 Э4 Э5			
1.7	Модель алгоритмов и метод обучения. Функционал качества /Ср/	3	4	ПК-4	Л1.1 Л1.2Л2.1Л3.2 Э2 Э3 Э4 Э5			
	Раздел 2. Метрические алгоритмы. Байесовский подход к обучению							
2.1	Метрические алгоритмы. Байесовский подход к обучению. Методы опорных векторов /Лек/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5	2		Интерактивное обсуждение метода опорных векторов
2.2	Практическая работа №2. Классификация изображений /Пр/	3	9	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5		9	
2.3	Алгоритм отбора эталонов. Проклятие размерности. /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
2.4	Выбор метрики. Применение формула Байеса к задаче машинного обучения. /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
2.5	Построение приближенной плотности распределения /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
2.6	Логистическая регрессия /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
2.7	Бинаризация признаков /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			

2.8	Скоринг. Смеси распределений /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
2.9	EM-алгоритм восстановления смеси /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
	Раздел 3. Линейные алгоритмы классификации и регрессия							
3.1	Линейные алгоритмы классификации. Метод опорных векторов. /Лек/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
3.2	Практическая работа №2. Классификация по нескольким меткам /Пр/	3	9	ПК-4	Л1.3 Л1.2Л2.1Л3.1 Э2 Э3	2	9	
3.3	Метод стохастического градиента /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
3.4	ROC и AUC /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
3.5	Нелинейное обобщение SVM /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
3.6	SVM-регрессия. L1 регуляризация /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
3.7	Практическая работа №4. Обучение современной модели классификации изображений /Пр/	3	9	ПК-4	Л1.3Л2.1Л3.2 Л3.1 Э1 Э2		9	Совместное использование Google CoLab
	Раздел 4. Логические методы классификации. Композиции алгоритмов							
4.1	Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация. /Лек/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5	2		Обсуждение проблем тематического моделирования
4.2	Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов /Лек/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
4.3	Непараметрическая регрессия /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
4.4	Критерий качества закономерностей. Поиск закономерностей /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
4.5	Алгоритмы классификации на основе логических закономерностей /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
4.6	Латентный семантический анализ (LSA) /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
4.7	Вероятностный LSA /Ср/	3	4	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			
4.8	Латентное размещение Дирихле (LDA). EM-алгоритм /Ср/	3	5,8	ПК-4	Л1.1Л2.1Л3.2 Э2 Э3 Э4 Э5			

4.9	/ЗачётСОц/	3		ПК-4	Л1.1 Л1.2 Л1.3Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4			
4.10	/КрТО/	3	0,2					

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Контрольные вопросы и задания

Вопросы для проверки уровня обученности ЗНАТЬ: 1. Основные определения: прецедент, обучающая выборка, признаки объектов, виды признаков, матрица объектов-признаков. Модель алгоритмов, метод обучения, функционал качества алгоритма. 2. Вероятностная постановка задачи обучения. Принцип максимума правдоподобия. Связь максимизации правдоподобия и минимизации эмпирического риска. Проблема переобучения и обобщающая способность алгоритма. Состоятельные методы обучения. Эмпирические оценки обобщающей способности. 3. Логистическая регрессия. Случайные величины с экспонентным законом распределения. Теорема о линейности байесовского классификатора (с доказательством). Бинаризация признаков. Скоринг. 4. Смеси распределений. EM-алгоритм разделения смеси. Смеси многомерных нормальных распределений. 5. Линейные алгоритмы классификации. Модель Мак Каллока-Питтса, алгоритм стохастического градиента для минимизации функционала среднего риска. Частные случаи. Сходимость метода СГ с правилом Хэбба с доказательством. Эвристики для улучшения сходимости и обобщающей способности. 6. Логические методы классификации. Понятие информативности предиката: эвристическое, вероятностное, энтропийное. Поиск информативных закономерностей. Построение решающего дерева и решающего дерева. Редукция деревьев. Применение деревьев для решения задачи регрессии. Небрежные решающие деревья. 7. Композиции алгоритмов. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг, метод случайных подпространств. Случайные лес. 8. Тематическое моделирование. Векторная модель текста, TF-IDF. Недостатки векторной модели. Тематические модели: LSA, PLSA. Распределение Дирихле. Тематическая модель LDA. 9. Задачи компьютерного зрения. Признаки изображений: глобальные, локальные. Применение сверточных нейронных сетей для построения признаков. Вопросы для проверки уровня обученности УМЕТЬ: 1. Выбор алгоритма для вероятностной постановки задачи. Функционал среднего риска. Метрические алгоритмы классификации. Обобщенный метрический классификатор. Виды особенности частных случаев: методы ближайшего соседа, k ближайших соседей, взвешенных соседей, парзеновского окна постоянной и переменной ширины. 2. Кривая ошибок ROC и AUC. Формула вычисления AUC. Градиентная максимизация AUC. 3. Метод опорных векторов (SVM). Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Функция Лагранжа. Классификация объектов в зависимости от значений множителей Лагранжа. Двойственная задача. Обучение SVM. Нелинейные обобщение SVM. SVM регрессия. Lasso SVM. 4. Ранжирование и рекомендательные системы. Постановка задачи. Оценки качества. Алгоритмы построения ранжирующих систем: поточечный, попарный и списочный. Их сильные и слабые стороны. Вопросы для проверки уровня обученности ВЛАДЕТЬ: 1. Классификация объектов по значению отступа. Алгоритм STOLP отбора эталонных объектов. 2. Выбор метрики и проклятие размерности. 3. Приближенное вычисление плотности распределения. Параметрический и непараметрический подходы. Наивный байесовский классификатор. Одномерный случай. Многомерный случай. Проблемы мультиколлинеарности и выбросов. 4. Алгоритмы восстановления регрессии. Метод наименьших квадратов. Многомерная линейная регрессия. Подход с использованием SVD-разложения матрицы. Гребневая регрессия. Метод главных компонент PCA. Непараметрическая регрессия. Проблема выбросов. Алгоритм LOWESS. 5. Кластеризация. Близость и связанность. EM-алгоритм, метод k-средних. DBSCAN. Выбор Eps и MinPts.

5.2. Темы курсовых работ (проектов)

не предусмотрены

5.3. Фонд оценочных средств

Примеры вопросов для контрольного опроса приведены в Приложении 2. Примеры заданий для практических работ приведены в Приложении 3.

Темы для рефератов приведены в Приложении 4.			
5.4. Перечень видов оценочных средств			
Контрольный опрос; Практические работы; Реферат. Виды шкал оценивания представлены в Приложении 1.			
6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)			
6.1. Рекомендуемая литература			
6.1.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Федин Ф.О.	Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие	М.: Московский городской педагогический университет 2014
Л1.2	Саймон Хайкин	Нейронные сети. Полный курс.: Учебник.	Вильямс 2016
Л1.3	Уэс Маккинли	Python и анализ данных: Электронный ресурс	Саратов: Профобразование 2017
6.1.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Федин Ф.О.	Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие	М.: Московский городской педагогический университет 2012
Л2.2	Э. Сигел	Практическая бизнес-статистика: Пер с англ.	Москва .: Издательский дом "Вильямс" 2002
6.1.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	сост. Истомина А.П.	Анализ данных качественных исследований: практикум	Ставрополь: Северо-Кавказский федеральный университет
Л3.2	Новоселов В.И.	Математическая логика и теория алгоритмов: Учебно-методическое пособие	Бишкек: Изд-во КРСУ 2011
6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"			
Э1	Неделько В.М. Основы статистических методов машинного обучения [Электронный ресурс] : учебное пособие / В.М. Неделько. — Электрон. текстовые данные. — Новосибирск: Новосибирский государственный технический университет, 2010. — 72 с. — 978-5-7782-1385-2. — Режим доступа: http://www.iprbookshop.ru/45418.html		http://www.iprbookshop.ru/45418.html
Э2	Уэс Маккинли Python и анализ данных [Электронный ресурс] / Маккинли Уэс. — Электрон. текстовые данные. — Саратов: Профобразование, 2017. — 482 с. — 978-5-4488-0046-7. — Режим доступа: http://www.iprbookshop.ru/64058.html		http://www.iprbookshop.ru/64058.html
Э3	Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс] : учебное пособие / С.В. Пальмов. — Электрон. текстовые данные. — Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 с. — 2227-8397. — Режим доступа: http://www.iprbookshop.ru/75376.html		http://www.iprbookshop.ru/75376.html
Э4	М. Тим Джонс Программирование искусственного интеллекта в приложениях [Электронный ресурс] / ТимДжонс М.. — Электрон. текстовые данные. — Саратов: Профобразование, 2017. — 310 с. — 978-5-4488-0116-7. — Режим доступа: http://www.iprbookshop.ru/63950.html		http://www.iprbookshop.ru/63950.html
Э5	Потапов А.С. Технологии искусственного интеллекта [Электронный ресурс] / А.С. Потапов. — Электрон. текстовые данные. — СПб. : Университет ИТМО, 2010. — 218 с. — 2227-8397. — Режим доступа: http://www.iprbookshop.ru/68201.html		http://www.iprbookshop.ru/68201.html
6.3. Перечень информационных и образовательных технологий			
6.3.1 Компетентностно-ориентированные образовательные технологии			
6.3.1.1	Изучение дисциплины студентами осуществляется в форме лекций, практических занятий в аудиторных условиях		
6.3.1.2	(компьютерные классы) и в процессе самостоятельной работы, контроля знаний.		
6.3.1.3	Теоретическая информация представляется в виде компьютерных презентаций с использованием мультимедийных		
6.3.1.4	средств.		
6.3.1.5	Практические занятия проводятся в компьютерных классах, оснащенных персональными компьютерами с		
6.3.1.6	необходимыми параметрами и с установленным необходимым программным обеспечением. Используется		
6.3.1.7	Интернет для получения дополнительной информации. Используется дискуссионный метод проведения занятий,		

6.3.1.8	где студенты могут высказать свое мнение по обсуждаемой проблеме.
6.3.2 Перечень информационных справочных систем и программного обеспечения	
6.3.2.1	http://www.machinelearning.ru/ коллекция материалов по машинному обучению
6.3.2.2	http://archive.ics.uci.edu/ml/ коллекция прикладных задач
6.3.2.3	https://www.python.org/ - дистрибутивы языка python
6.3.2.4	http://scikit-learn.org/stable/modules/classes.html - описание пакета sklearn
6.3.2.5	Программное обеспечение компьютеров учебного класса:
6.3.2.6	1. Операционная система Microsoft Windows (XP, Vista, Win7, Win8) или
6.3.2.7	Linux
6.3.2.8	2. Microsoft Office (коммерческое ПО) или Open Office (бесплатное ПО)
6.3.2.9	3. Python 2 или Python 3 с основными библиотеками.
7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)	
7.1	Для проведения курса требуются следующие технические ресурсы:
7.2	1. Терминальный класс с компьютерами на базе процессоров не хуже PIV 3Hz, 1Gb RAM;
7.3	2. Проектор.
8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)	
<p>Технологическая карта дисциплины приведена в Приложении 5. В процессе обучения рекомендуется обратить особое внимание на практические занятия и работу над практическим заданием. Цель – решить задачу не самым простым для реализации образом, а подобрать оптимальную модель, выбрать в ее рамках алгоритм и настроить его, после чего с помощью различных методик оценки качества решения продемонстрировать его преимущество. Для этого требуется хорошо разбираться в теоретической части курса и в возможностях пакета программ sklearn. Также полезно будет внимательно слушать материалы, докладываемые другими студентами, так как они дают более широкое представление о курсе и затрагивают моменты, не попавшие в лекции.</p>	

ПРИЛОЖЕНИЕ 1
ШКАЛА ОЦЕНИВАНИЯ РЕФЕРАТА (рубежный контроль)

№	Наименование показателя	Отметка (в %)
1	Во введении четко сформулирован тезис, соответствующий теме реферата, выполнена задача заинтересовать читателя	85 - 100
2	Деление текста на введение, основную часть и заключение	
3	В основной части логично, связно и полно доказывается выдвинутый тезис	
4	Заключение содержит выводы, логично вытекающие из содержания основной части	
5	Все требования, предъявляемые к заданию, выполнены	
6	При защите реферата демонстрирует полное понимание проблемы и для выражения своих мыслей не пользуется упрощенно-примитивным языком.	
1	Во введении четко сформулирован тезис, соответствующий теме реферата, в известной мере выполнена задача заинтересовать читателя	75 – 84
2	В основной части логично, связно, но не достаточно полно доказывается выдвинутый тезис	
3	Заключение содержит выводы, логично вытекающее из содержания основной части	
4	При защите реферата демонстрирует понимание проблемы и для выражения своих мыслей не пользуется упрощенно-примитивным языком.	
1	Во введении тезис сформулирован не четко и не вполне соответствует теме реферата	60 - 74
2	В основной части выдвинутый тезис доказывается недостаточно логично (убедительно) и последовательно	
3	Заключенные выводы не полностью соответствуют содержанию основной части	
4	При защите реферата демонстрирует не полное понимание проблемы и язык работы в целом не соответствует требуемому уровню	
1	Во введении тезис отсутствует или не соответствует теме реферата	40 - 59
2	Деление текста на введение, основную часть и заключение отсутствует	
3	В основной части нет логичного последовательного раскрытия темы	
4	Выводы не вытекают из основной части	
5	При защите реферата демонстрирует полное непонимание проблемы и язык работы можно оценить, как «примитивный».	
1	Работа написана не по теме	менее 58

ШКАЛА ОЦЕНИВАНИЯ ПРАКТИЧЕСКИХ (ЛАБОРАТОРНЫХ) РАБОТ (текущий контроль)

- 85-100 % - Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию выполнены.
- 70-84 % - Демонстрирует значительное понимание проблемы. Все требования, предъявляемые к заданию выполнены.
- 60-69 % - Демонстрирует частичное понимание проблемы. Большинство требований, предъявляемых к заданию выполнены.
- 31-60 % - Демонстрирует небольшое понимание проблемы. Многие требования, предъявляемые к заданию не выполнены.

- 0-30 % - Демонстрирует непонимание проблемы и даже не было попытки решить задачу.

ШКАЛА ОЦЕНИВАНИЯ КОНТРОЛЬНЫХ РАБОТ (рубежный контроль)

- 85-100 % - Демонстрирует полное понимание проблемы. Все задания выполнены.
- 70-84 % - Демонстрирует значительное понимание проблемы. Все задания выполнены, но содержат некоторые неточности.
- 60-69 % - Демонстрирует частичное понимание проблемы. Большинство требований, предъявляемых к заданию, выполнены.
- 31-60 % - Демонстрирует небольшое понимание проблемы. Многие требования, предъявляемые к заданию, не выполнены.
- 0-30 % - Демонстрирует непонимание проблемы или нет ответа и даже не было попытки решить задачу.

ШКАЛА ОЦЕНИВАНИЯ ПИСЬМЕННОГО ОПРОСА

(промежуточный контроль – «ЗНАТЬ»)

Отметкой (7-10- баллов) оценивается ответ, который показывает прочные знания теоретических основ дисциплины, понимание и правильное применение терминологии, правильные ответы на 75-100% вопросов

Отметкой (5-7 баллов) оценивается ответ, который показывает знание теоретических основ дисциплины, но неполное понимание и не всегда правильное применение терминологии, даны правильные ответы на 50-74% вопросов, в ответах допущено некоторое количество неточностей.

Отметкой (3-4 баллов) оценивается ответ, свидетельствующий о знакомстве с некоторыми теоретическими основами дисциплины. Даны правильные ответы на 25-49% вопросов, допущены неточности и ошибки.

Отметкой (2 балла) оценивается ответ, обнаруживающий незнание теоретических основ дисциплины. Отмечается отсутствие логичности и последовательности в ответе. Менее 25% правильных ответов. Допущены серьезные ошибки в содержании ответа.

Отметкой (0-1 балл) оценивается ответ, при котором студент демонстрирует непонимание поставленных вопросов, или нет ответа.

ШКАЛА ОЦЕНИВАНИЯ ПРАКТИЧЕСКИХ ЗАДАНИЙ

(промежуточный контроль – «УМЕТЬ и ВЛАДЕТЬ»)

Отметкой (8-10 баллов) оценивается ответ, при котором студент правильно отвечает на поставленные вопросы, Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию, выполнены.

Отметкой (5-7 баллов) оценивается ответ, при котором студент в основном правильно отвечает на поставленные вопросы. Демонстрирует значительное понимание проблемы. Большинство требований, предъявляемых к заданию, выполнены.

Отметкой (2-4 баллов) оценивается ответ, при котором студент в основном не правильно отвечает на поставленные вопросы. Демонстрирует частичное или небольшое понимание проблемы. Многие требования, предъявляемые к заданию, не выполнены.

Отметкой (0 -1 балл) оценивается ответ, при котором студент демонстрирует непонимание проблемы или нет ответа и даже не было попытки решить задачи.

Вопросы для контрольного опроса

Контрольный опрос по заданной тематике проводится с целью рубежного контроля.

Примерные вопросы выглядят следующим образом:

Контрольный опрос №1

Как на компьютере отображается изображение в оттенках серого? А цветное?

Как структурированы файлы и каталоги в датасете NIST_SAMPLE?

Почему- именно так?

Объясните, как работает подход «пиксельного сходства» при классификации цифр.

Что такое генератор списка? Создайте такой, который будет выбирать нечетные числа из списка и удваивать их.

Что такое тензор ранга 3?

В чем разница между рангом тензора и его формой? Как можно получить ранг, исходя из формы?

Что такое RMSE и норма L1?

Как можно применить вычисление к тысячам чисел сразу и в несколько тысяч раз быстрее, чем при использовании цикла Python?

Создайте тензор 3X3 или массив, содержащий числа от 1 до 9. Удвойте его. Выберите четыре числа из правой нижней части.

Что такое транслирование?

Как обычно вычисляются метрики: с помощью обучающей выборки или контрольной?

Почему?

Что такое SGD?

Почему SGD использует мини-пакеты?

Из каких семи шагов состоит SGD для машинного обучения?

Как мы инициализируем веса модели?

Что такое потери?

Почему не всегда можно использовать высокую скорость обучения?

Что такое градиент?

Нужно ли вам знать, как вычислять градиенты самостоятельно?

Почему мы не можем использовать в качестве функции потерь точность?

Нарисуйте сигмоидную функцию. В чем особенность ее формы?

В чем отличие между функцией потерь и метрикой?

Какая функция вычисляет новые веса на основе скорости обучения?

Что делает класс DataLoader?

Напишите псевдокод, отображающий основные шаги, совершаемые в каждой эпохе для SGD.

Создайте функцию, которая при получении двух аргументов [1,2,3,4] и 'abcd' будет возвращать [(1, 'a'), (2, 'b'), (3, 'c'), (4, 'd')]. В чем особенность структуры полученных на выводе данных?

Какова роль view в PyTorch?

Что такое параметры смещения в нейронной сети? Зачем они нам нужны?

Каково назначение оператора @ в Python?

Что делает метод backward?

Зачем нужно обнулять градиенты?

Какую информацию нужно передавать в Learner?

Покажите код Python или псевдокод для основных шагов цикла обучения.

Что такое ReLU? Нарисуйте ее график для значений от -2 до $+2$.

Что такое функция активации?

В чем разница между `F.relu` и `nn.ReLU`?

Теорема универсальной аппроксимации показывает, что любую функцию можно аппроксимировать настолько, насколько необходимо, используя всего одну нелинейность. Зачем же мы обычно используем больше?

Контрольный опрос №2. Классификация изображений

Почему мы сначала увеличиваем размер с помощью CPU, а затем уменьшаем его на GPU?

Если вы не знакомы с регулярными выражениями, найдите по ним руководство и несколько наборов задач для решения. На сайте книги предложено несколько вариантов.

Какими двумя способами данные чаще всего предоставляются для большинства датасетов глубокого обучения?

Изучите документацию к классу `L` и попробуйте использовать несколько из новых методов, которые он добавляет.

Ознакомьтесь с документацией для Python-модуля `pathlib` и попробуйте использовать несколько методов класса `Path`.

Приведите два примера того, как преобразования изображений могут ухудшить качество данных.

Какой метод в `fastai` позволяет просматривать данные в `DataLoaders`?

Какой метод в `fastai` позволяет вам отлаживать `DataBlock`?

Стоит ли вам отложить обучение модели до того, как данные будут тщательно очищены?

Из каких двух частей составляется функция перекрестной энтропии

PyTorch?

Какие два свойства активаций обеспечивает `softmax`? Почему это важно?

В каких случаях вам может потребоваться отсутствие у активаций этих двух свойств?

Вычислите столбцы `exp` и `softmax` на рис. 5.3 самостоятельно (то есть

электронной таблице, калькуляторе или блокноте).

Почему нельзя использовать `torch.where` при создании функции потерь для датасетов, в которых целевая метка может иметь более двух категорий?

Каково значение $\log(-2)$? Почему?

Контрольный опрос № 3

Каковы два основных рекомендуемых варианта выбора скорости обучения- при использовании ее искателя?

Какие два шага выполняет метод `fine_tune`?

Как получить исходный код метода или функции в Jupyter Notebook?

Что такое дискриминативные скорости обучения?

Как Python-объект `slice` интерпретируется при передаче в качестве скорости обучения в `fastai`?

Почему обучение с остановкой считается плохим выбором при использовании подхода `1cycle`?

В чем разница между `resnet50` и `resnet101`?

Для чего используется `to_fp16`?

Как можно использовать классификацию по нескольким меткам для повышения удобства и качества работы классификатора медведей?

Как мы кодируем зависимую переменную в задаче классификации по нескольким меткам?

Как обращаться к строкам и столбцам `DataFrame`, если рассматривать его как матрицу?

Как извлечь из `DataFrame` столбец по его имени?

В чем отличие `Dataset` от `DataLoader`?

Что обычно содержит объект `Datasets`?

Что обычно содержит объект `DataLoaders`?

Каково назначение `lambda` в Python?

С помощью каких методов настраивается создание независимых и зависимых переменных с помощью API блока данных?

Почему `softmax` не подходит в качестве функции активации при использовании цели, закодированной унитарно?

Почему `nll_loss` не подходит в качестве функции потерь при использовании цели, закодированной унитарно?

В чем разница между `nn.BCELoss` и `nn.BCEWithLogitsLoss`?

Почему мы не можем использовать обычную метрику точности в задаче классификации по нескольким меткам?

Когда допускается настройка параметра с помощью контрольной выборки?

Как в `fastai` реализуется `y_range`? (Попробуйте реализовать ее самостоятельно и протестировать не подглядывая.)

Что подразумевает задача по регрессии? Какую функцию потерь следует для нее использовать?

Контрольный опрос №4

В чем отличие между `ImageNet` и `Imagenette`? С какой базой данных и в каких ситуациях следует экспериментировать?

Что такое нормализация?

Почему нас не волновала нормализация, когда мы использовали предварительно обученную модель?

Что такое постепенное увеличение размера?

Реализуйте постепенное увеличение размера в собственном проекте.

Помогло- ли оно?

Что такое аугментация во время тестирования? Как она используется в fastai?

Использование ГТА в процессе вывода модели замедляет или ускоряет этот процесс?

Что такое Mixup? Как этот метод используется в fastai?

Почему Mixup не позволяет модели быть слишком уверенной?

Почему обучение с Mixup в течение пяти эпох приводит к худшему показателю эффективности, чем обучение без него?

Какая идея стоит за сглаживанием меток?

Какие проблемы с данными может решить сглаживание меток?

Какая цель будет ассоциироваться с индексом 1 при использовании сглаживания меток по пяти категориям?

С чего следует начинать подготовку к проведению быстрых экспериментов по прототипированию с новым датасетом?

ПРАКТИЧЕСКИЕ РАБОТЫ

Практика 1. Обучение классификатора цифр***Задания***

Создайте собственную реализацию Learner с нуля, взяв за основу обучающий цикл, показанный в этой лекции.

Выполните все шаги этой работы, используя полные датасеты MNIST (для всех цифр, а не только для троек и семерок). Это значимый проект, который потребует от вас немало времени. Вам понадобится проделать самостоятельное исследование, чтобы понять, как преодолеть те или иные возникающие преграды.

Практика 2. Классификация изображений***Задания***

Прочтите руководство по Pandas DataFrames и поэкспериментируйте с несколькими методами, которые покажутся вам наиболее интересными. Рекомендованные учебные материалы доступны на сайте курса.

Обучите повторно классификатор медведей с помощью классификации по нескольким меткам. Попробуйте добиться от него эффективной работы с изображениями, где нет медведей, включая показ соответствующей информации в веб-приложении. Попробуйте изображение с двумя видами медведей. Проверьте, изменилась ли точность из-за использования классификации по нескольким меткам для датасета с одиночными метками.

Практика 3. Классификация по нескольким меткам***Задания***

1. Прочтите руководство по Pandas DataFrames и поэкспериментируйте с несколькими методами, которые покажутся вам наиболее интересными.
-Рекомендованные учебные материалы доступны на сайте книги.
2. Обучите повторно классификатор медведей с помощью классификации по нескольким меткам. Попробуйте добиться от него эффективной работы с изображениями, где нет медведей, включая показ соответствующей информации в веб-приложении. Попробуйте изображение с двумя вида-

ми медведей. Проверьте, изменилась ли точность из-за использования классификации по нескольким меткам для датасета с одиночными метками.

Практическа работа №4. Обучение современной модели классификации изображений

Задания

Используйте документацию `fastai` для построения функции, обрезающей изображение до квадрата в каждом из четырех углов. Затем реализуйте метод ТГА, усредняющий прогнозы обрезки по центру и тех четырех частей. Улучшился ли результат? Лучше ли этот метод, чем ТГА в `fastai`?

Найдите и прочитайте работу по исследованию `Mixup` на `arXiv`. Ознакомьтесь с одной или двумя недавними статьями, описывающими вариации `Mixup`, и попробуйте реализовать эти вариации в своем проекте.

Найдите скрипт обучения `Imagenette` с использованием `Mixup` и примените его в качестве примера для построения скрипта длительного обучения в своем проекте. Выполните его и посмотрите, улучшит ли это результат.

Прочтите сноску «Сглаживание меток, научная работа», затем откройте связанный с ней раздел оригинала работы и попробуйте его понять. Не стесняйтесь обращаться за помощью!

ПРИМЕРНЫЕ ТЕМЫ РЕФЕРАТОВ

Написание реферата по заданной тематике и представление его перед группой проводится с целью текущего контроля. Примерные темы рефератов выглядят следующим образом :

1. Система полигон сравнения алгоритмов распознавания.
2. Современные технологии выбора признаков
3. Динамическое распознавание
4. Бикластеризация
5. Метод независимых компонент.
6. Задача частичного обучения. Алгоритм Co-learning.
7. Деревья классификаций. Алгоритмы CART, C4.5.
8. Парзеновские окна.
9. Методы прогнозирования временных рядов
10. Линейный дискриминант Фишера
11. Визуализация данных
12. Композиция алгоритмов.
13. Ассоциативные правила. Алгоритм APRIORY

Технологическая карта дисциплины **Машинное обучение и анализ данных**

2 семестр

85 – 100 баллов– «отлично»

70 – 84 баллов– «хорошо»

60-69 баллов– «удовлетворительно»

менее 60 баллов– «неудовлетворительно»

Название модулей дисциплины согласно РПД	Контроль	Форма контроля	Формы контроля		График контроля
			зачетный минимум	зачетный максимум	
1. Введение в машинное обучение	текущий	Сдача практической работы №1	5	8	5
	рубежный	Контрольный опрос №1	5	8	5
2. Метрические алгоритмы. Байесовский подход к обучению	текущий	Сдача практической работы №2	5	9	9
	рубежный	Контрольный опрос №2	5	9	9
3. Линейные алгоритмы классификации и регрессия	текущий	Сдача практической работы №3,4	5	9	13
	рубежный	Контрольный опрос №3	5	9	13
4. Логические методы классификации. Композиции алгоритмов	текущий	Сдача реферата	5	9	17
	рубежный	Контрольный опрос №4	5	9	17
ВСЕГО за семестр			40	70	19
Промежуточный контроль (Экзамен)			20	30	
Семестровый рейтинг по дисциплине			60	100	