

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова»
Институт математики, физики и компьютерных наук
Кафедра информационных систем и методов искусственного интеллекта

Утверждена на заседании
Ученого совета ИМФКН
«___» _____ 20__ г.
Протокол №__

Рабочая программа дисциплины

Прикладные задачи машинного обучения

Направление подготовки
02.04.03 Математическое обеспечение и администрирование
информационных систем

Квалификация
Магистр

Форма обучения
Очная

Улан-Удэ
2025

Пояснительная записка

Цели освоения дисциплины

Цель дисциплины - освоение современных методов и технологий решения прикладных задач машинного обучения

Место дисциплины в структуре образовательной программы

Дисциплина изучается в 3 семестре и является обязательной, входит в обязательную часть блока Б1 ОП по направлению подготовки 02.04.03 «Математическое обеспечение и администрирование информационных систем» (магистратура). Основывается на курсах: "Методы и технологии глубокого обучения", "Методы оптимизации".

Планируемые результаты обучения по дисциплине и индикаторы достижения компетенций.

В результате освоения дисциплины студент должен:

Знать:

- определения основных понятий, содержательную структуру и типизацию прикладных задач машинного обучения;
- основные методологические подходы, принципы и алгоритмы построения и обучения различных моделей прикладных задач;
- комплекс эвристик для повышения качества моделей.

Уметь:

- формализовывать прикладные задачи машинного обучения;
- выбирать в соответствии условиям задачи модели, методы и технологии ее решения;
- решать прикладные задачи машинного обучения с применением технологий программирования

Владеть:

- современным понятийным аппаратом машинного обучения, в т.ч. глубокого обучения;
- современными инструментальными средствами и технологиями программной реализации решения прикладных задач.

Планируемые результаты освоения образовательной программы:

- | | |
|--------|--|
| ПК-2 | Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач |
| ПК-2.2 | Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий |
| ПК-2.3 | Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области |
| ПК-3 | Способен руководить проектами по созданию комплексных систем искусственного интеллекта |
| ПК-3.1 | Применяет методы и средства разработки систем искусственного интеллекта |
| ПК-3.2 | Руководит разработкой архитектуры комплексных систем искусственного интеллекта |

Объем дисциплины в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 5 зачетные единицы, 180 часа.

№ Название разделов дисциплины	Лекция	Лабораторная работа	Самостоятельная работа
Семестр 3	16	30	98
1 Прикладные нейросетевые модели	4	8	38
2 Специальные прикладные задачи машинного обучения	12	22	60

Тематическое планирование курса

Темы

Прикладные нейросетевые модели

Семестр 3

Векторные представления текстов и графов

Лекция. 2(0) ч. Векторные представления текста. Гипотеза дистрибутивной семантики. Модели CBOW и SGNS из программы word2vec. Иерархический SoftMax. Модель FastText. Векторные представления графов. Многомерное шкалирование (multidimensional scaling, MDS). Векторное представление соседства (stochastic neighbor embedding, SNE и tSNE). Матричные разложения (graph factorization). Модели случайных блужданий DeepWalk, node2vec. Обобщённый автокодировщик на графах GraphEDM. Представление о графовых нейронных сетях (graph neural network, GNN). Передача сообщений по графу (message passing).

Лабораторная работа. 4(0) ч. Применение моделей CBOW и SGNS из программы word2vec

Самостоятельная работа. 18(0) ч. Модели CBOW и SGNS из программы word2vec. Иерархический SoftMax. Векторное представление соседства (stochastic neighbor embedding, SNE и tSNE).

Модели внимания и трансформеры

Лекция. 2(0) ч. Задачи обработки и преобразования последовательностей (sequence to sequence). Рекуррентная сеть с моделью внимания. Разновидности моделей внимания: многомерное, иерархическое, Query–Key–Value, внутреннее (self-attention). Модели внимания на графах (Graph Attention Network). Задача классификации вершин графа. Трансформеры. Особенности архитектуры кодировщика и декодировщика. Критерии обучения и оценивание качества (предобучение). Модель BERT. Прикладные задачи: машинный перевод, аннотирование изображений. Модели внимания и трансформеры для текстов, изображений, графов.

Лабораторная работа. 4(0) ч. Прикладные задачи: машинный перевод, аннотирование изображений. Модели внимания и трансформеры для текстов, изображений, графов.

Самостоятельная работа. 20(0) ч. Рекуррентная сеть с моделью внимания. Разновидности моделей внимания: многомерное, иерархическое, Query–Key–Value,

внутреннее (self-attention). Трансформеры. Особенности архитектуры кодировщика и декодировщика. Модель BERT. Прикладные задачи: машинный перевод, аннотирование изображений. Модели внимания и трансформеры для текстов, изображений, графов.

Специальные прикладные задачи машинного обучения

Семестр 3

Тематическое моделирование

Лекция. 2(0) ч. Задача тематического моделирования коллекции текстовых документов. Метод максимума правдоподобия. Лемма о максимизации гладкой функции на симплексах (применение условий Каруша–Куна–Таккера). Аддитивная регуляризация тематических моделей. Регуляризованный EM-алгоритм, теорема о стационарной точке. Элементарная интерпретация EM-алгоритма. Вероятностный латентный семантический анализ PLSA. EM-алгоритм. Латентное размещение Дирихле LDA. Метод максимума апостериорной вероятности. Сглаженная частотная оценка условной вероятности. Небайесовская интерпретация LDA. Регуляризаторы разреживания, сглаживания, частичного обучения, декоррелирования. Мультиязычная тематическая модель. Мультиязычная тематическая модель. Регуляризаторы классификации и регрессии. Модель битермов WNTM. Модель связанных документов. Иерархическая тематическая модель.

Лабораторная работа. 4(0) ч. Пример задачи тематического моделирования коллекции текстовых документов

Самостоятельная работа. 10(0) ч. Вероятностный латентный семантический анализ PLSA. EM-алгоритм. Латентное размещение Дирихле LDA. Метод максимума апостериорной вероятности. Сглаженная частотная оценка условной вероятности. Небайесовская интерпретация LDA. Регуляризаторы разреживания, сглаживания, частичного обучения, декоррелирования.

Обучение ранжированию

Лекция. 2(0) ч. Постановка задачи обучения ранжированию. Примеры. Поточечные методы Ранговая регрессия. Ранговая классификация, OC-SVM. Парные методы: RankingSVM, RankNet, LambdaRank. Списочные методы. Признаки в задаче ранжирования поисковой выдачи: текстовые, ссылочные, кликовые. TF-IDF, Okapi BM25, PageRank. Критерии качества ранжирования: Precision, MAP, AUC, DCG, NDCG, pFound. Глубокая структурированная семантическая модель DSSM (Deep Structured Semantic Model).

Лабораторная работа. 4(0) ч. Обзор: поточечные методы, парные методы (RankingSVM, RankNet, LambdaRank), списочные методы.

Самостоятельная работа. 10(0) ч. Поточечные методы Ранговая регрессия. Ранговая классификация, OC-SVM. Парные методы: RankingSVM, RankNet, LambdaRank.

Рекомендательные системы

Лекция. 2(0) ч. Задачи коллаборативной фильтрации, транзакционные данные. Корреляционные методы user-based, item-based. Задача восстановления пропущенных значений. Меры сходства. Разреженная линейная модель (Sparse Linear Method, SLIM). Латентные методы на основе матричных разложений. Метод главных компонент для разреженных данных (LFM, Latent Factor Model). Метод стохастического градиента. Неотрицательные матричные разложения NMF. Метод чередующихся наименьших квадратов ALS. Вероятностный латентный семантический анализ PLSA. Модель с

учётом неявной информации (implicit feedback). Автокодировщики для коллаборативной фильтрации. Учёт дополнительных признаков данных в матричных разложениях и автокодировщиках. Линейная и квадратичная регрессионные модели, libFM. Гиперграфовая транзакционная тематическая модель для учёта дополнительных данных. Измерение качества рекомендаций. Меры разнообразия (diversity), новизны (novelty), покрытия (coverage), догадливости (serendipity).

Лабораторная работа. 4(0) ч. Обзор прикладных задач и технологий решения задач коллаборативной фильтрации

Самостоятельная работа. 10(0) ч. Латентные методы на основе матричных разложений. Метод главных компонент для разреженных данных (LFM, Latent Factor Model). Метод стохастического градиента. Неотрицательные матричные разложения NMF. Метод чередующихся наименьших квадратов ALS. Вероятностный латентный семантический анализ PLSA.

Поиск ассоциативных правил

Лекция. 2(0) ч. Понятие ассоциативного правила и его связь с понятием логической закономерности. Примеры прикладных задач: анализ рыночных корзин, выделение терминов и тематики текстов. Алгоритм APriori. Два этапа: поиск частых наборов и рекурсивное порождение ассоциативных правил. Недостатки и пути усовершенствования алгоритма APriori. Алгоритм FP-growth. Понятия FP-дерева и условного FP-дерева. Два этапа поиска частых наборов в FP-growth: построение FP-дерева и рекурсивное порождение частых наборов. Общее представление о динамических и иерархических методах поиска ассоциативных правил. Алгоритм TopMine для поиска коллокаций и терминов в текстах.

Лабораторная работа. 4(0) ч. Анализ алгоритма APriori. Два этапа: поиск частых наборов и рекурсивное порождение ассоциативных правил. Недостатки и пути усовершенствования алгоритма APriori.

Самостоятельная работа. 10(0) ч. Алгоритм APriori. Два этапа: поиск частых наборов и рекурсивное порождение ассоциативных правил. Недостатки и пути усовершенствования алгоритма APriori. Алгоритм FP-growth. Понятия FP-дерева и условного FP-дерева. Два этапа поиска частых наборов в FP-growth: построение FP-дерева и рекурсивное порождение частых наборов.

Инкрементное и онлайнное обучение

Лекция. 2(0) ч. Задачи инкрементного и онлайнного обучения. Оценивание инкрементного обучения. Кривые обучения. Ленивое обучение (метрические и непараметрические методы). Онлайнный отбор эталонных объектов. Онлайнный наивный байесовский классификатор. Онлайнный градиентный спуск OGD. Алгоритм Perceptron. Алгоритм Passive-Aggressive. Инкрементные решающие деревья ID5R. Рекуррентный метод наименьших квадратов RLS. Формула Шермана-Моррисона. Задача прогнозирования временных рядов. Эконометрические временные ряды с трендом и сезонностью. Экспоненциальное скользящее среднее. Модель Хольта. Модель Тейла-Вейджа. Модель Хольта-Уинтерса. Адаптивная селективная модель. Адаптивная композиция моделей. Онлайнное обучение ансамбля. Алгоритм Hedge, его свойства и интерпретация в задаче портфельного инвестирования. Онлайнное глубокое обучение. Алгоритм Hedge BackProp.

Лабораторная работа. 4(0) ч. Обзор программных технологий инкрементного и онлайнного обучения

Самостоятельная работа. 10(0) ч. Онлайновый градиентный спуск OGD. Алгоритм Perceptron. Алгоритм Passive-Aggressive. Инкрементные решающие деревья ID5R. Рекуррентный метод наименьших квадратов RLS. Формула Шермана-Моррисона.

Обучение с подкреплением

Лекция. 2(0) ч. Задача о многоруком бандите. Жадные и эpsilon-жадные стратегии. Метод UCB (upper confidence bound). Адаптивные стратегии на основе скользящих средних. Метод сравнения с подкреплением. Метод преследования. Постановка задачи в случае, когда агент влияет на среду. Ценность состояния среды. Ценность действия. Жадные стратегии максимизации ценности. Уравнения оптимальности Беллмана. Метод SARSA. Метод Q-обучения. Типизация методов на on-policy и off-policy. Глубокое Q-обучение нейронной сети DQN на примере обучения играм Atari. Градиентная оптимизация стратегии (policy gradient). Связь с максимизацией log-правдоподобия. Модели актор-критик. Модели с непрерывным управлением. Постановка задачи при моделировании среды. Типизация методов на model-free и model-based. Контекстный многорукий бандит. Линейная регрессионная модель с верхней доверительной оценкой LinUCB. Оценивание новой стратегии по большим историческим данным, сформированным при старых стратегиях

Лабораторная работа. 2(0) ч. Анализ адаптивных стратегии на основе скользящих средних. Метод сравнения с подкреплением. Метод преследования.

Самостоятельная работа. 10(0) ч. Жадные стратегии максимизации ценности. Уравнения оптимальности Беллмана. Метод SARSA. Метод Q-обучения. Типизация методов на on-policy и off-policy. Глубокое Q-обучение нейронной сети DQN на примере обучения играм Atari. Градиентная оптимизация стратегии (policy gradient). Связь с максимизацией log-правдоподобия.

БРС

Семестр	Контрольные точки	Баллы
3	Текущий контроль в разделе «Прикладные нейросетевые модели»	
	Лабораторная работа	20
	Лабораторная работа	10
3	Текущий контроль в разделе «Специальные прикладные задачи машинного обучения »	
	Лабораторная работа	10
	Лабораторная работа	20
3	Экзамен	
	Теоретические вопросы	40

Итого за семестр 3: 100

Учебно-методическое и информационное обеспечение учебного процесса

Образовательные технологии (в том числе на занятиях, проводимых в интерактивных формах).

Периодически используется технология проблемного обучения. Студентам даются сырые статистические данные из некоторой прикладной области (психология, социология, медицина, экономика и т.д.). Преподаватель формулирует задачу с точки зрения предметной области. Студенту необходимо правильно формализовать задачу и выбрать соответствующий метод анализа, затем решить ее с помощью специализированного программного пакета. Полученный результат необходимо интерпретировать с точки зрения предметной области.

Лекционные и лабораторные занятия проводятся с использованием презентаций.

Учебно-методические материалы, в том числе методические указания для обучающихся по освоению дисциплины

Методические указания к лекционным занятиям.

1. Если во время лекции все же не совсем разобрались в отдельных моментах раскрываемой темы, рекомендуется в индивидуальном порядке уточнить непонятные разделы у преподавателя во время лекции (поднять руку и задать вопрос), либо после нее. Скромность - качество замечательное, но в отдельных случаях быть скромным просто неразумно.
2. Для того, чтобы составлять качественные конспекты лекций, важно понять, что конспект – не дословно записанная речь преподавателя. Преподаватель вообще не обязан диктовать текст лекции под запись, он ведет согласно плану. Таким образом, в течение лекции студент тратит большую часть времени на восприятие информации, меньшую его часть – на ее запись.
3. Для повышения эффективности конспектирования материала рекомендуется воспользоваться следующими рекомендациями: 1) Убирайте только середину слова, а не середину и окончание (например, удачный «эф-ть», не удачный «эф.»). 2) В процессе лекции пишите часть слова, затем в тексте оставляйте место для второй его части, а на перерыве или после занятий (пока не забыли, о чем шла речь) вписывайте оставшуюся часть слова.
- 3) Заменяйте длинные русские слова короткими иностранными, например, несколько – some, выигрывать – win, использовать – use, экономический – economic и т.д.

Методические указания к лабораторным (практическим) занятиям.

1. В ходе лабораторных занятий обучающиеся фактически впервые сталкиваются с самостоятельной практической деятельностью в конкретной области – содействует становлению студентов как будущих специалистов. Поэтому, необходимо студенту проявить здесь особое усердие и получить ощутимый результат.
2. Результаты выполнения лабораторных (практических) работ нужно оформить в виде отчета. Как правило, отчет состоит из 3-х частей: план отчета (общая структура задания); расчетные формулы, блок-схема алгоритма, принципиальная часть программного кода, применяемые методы и средства (библиотеки, модули, структуры данных, службы, шаблоны классов, математические методы ит.п.), авторский проект решения задачи; выводы.
3. Перед сдачей лабораторных работ (практических заданий) необходимо повторить теоретический материал для более глубокого понимания и грамотного комментирования

выполненной работы преподавателю.

Методические указания к самостоятельной работе студента.

1. Выполняйте внеаудиторное задание в день его получения, а накануне занятия повторите его.
2. Для успешного выполнения задания создайте условия, которые отвечают требованиям гигиены умственного труда: удобное место, достаточное освещение, тишина, перерывы, необходимое оборудование.
3. Начинайте выполнять задание с его осмысления: определите цель, содержание, степень новизны, уровень усвоения, объем, сроки, этапы и приемы выполнения. Спланируйте и соблюдайте затем последовательность действий. Познакомьтесь с алгоритмом и эталоном выполнения задания.
4. Изучите вначале теоретическую основу задания (закон, правило, первоисточник и др.), затем принимайтесь за практическую работу.
5. Старайтесь выполнять задание самостоятельно, применяя знания и умения, усвоенные ранее.
6. Определите свой оптимальный ритм и режим работы.
7. Помните, что следование рекомендациям научной организации учебного труда экономит время, способствует достижению наилучших результатов.

Оценочные средства

По данной дисциплине разработаны оценочные средства, критерии их оценивания, а также методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций (в приложении).

Список литературы

Перечень основной и дополнительной литературы, необходимой для освоения дисциплины.

Основная

1. Нейронные сети: учеб. пособие/Горожанина Е. И.. —Самара: Изд-во ПГУТИ, 2017. —84 с. (Электронный ресурс ИРБИС")
Режим доступа: <https://lib.rucont.ru/efd/641652>
2. Манга: Машинное обучение/Араки М.. —Москва: ДМК Пресс, 2020. —214 с.
Режим доступа: <https://e.lanbook.com/book/179473>
3. Рекомендательные системы на практике/Фальк К.. —Москва: ДМК Пресс, 2020. —448 с.
Режим доступа: <https://e.lanbook.com/book/179458>
4. Компьютерное зрение. Передовые методы и глубокое обучение/Тёрк М.,Дэвис Р.. —Москва: ДМК Пресс, 2022. —690 с.
Режим доступа: <https://e.lanbook.com/book/314900>
5. Глубокое обучение с R и Keras/Шолле Ф.. —Москва: ДМК Пресс, 2023. —646 с.
Режим доступа: <https://e.lanbook.com/book/315488>

Дополнительная

1. Яхьяева Г. Э. Нечеткие множества и нейронные сети: учеб. пособие/Г. Э. Яхьяева. —М.: Интернет-Университет Информационных Технологий, 2008. —315 с.

2. Вероятностное машинное обучение. Введение/Мэрфи К. П.. —Москва: ДМК Пресс, 2022. —940 с.
Режим доступа: <https://e.lanbook.com/book/314891>
3. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных: научно-популярная литература/П. Флах ; пер. с англ. А. А. Слинкин. —Москва: ДМК Пресс, 2015. —400 с.
Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=69955

Перечень ресурсов информационно-коммуникационной сети «Интернет», необходимых для освоения дисциплины

Машинное обучение (курс лекций, К.В.Воронцов): <https://bit.ly/ML-Vorontsov>

Уроки PYTHON для начинающих:

https://www.youtube.com/playlist?list=PLA0M1Bcd0w8xIdFNA95aQrwJ_GQJEV8ko

Python для анализа данных: https://www.youtube.com/playlist?list=PLJb04_psp36AfoU-VWozLueaoD_GPZV5

Tensorflow 2.x - уроки (для Python):

<https://www.youtube.com/playlist?list=PLA0M1Bcd0w8ynD1umfubKq1OBYRXhXkmH>

Нейросети на Python: https://www.youtube.com/playlist?list=PLtPJ9IKvJ4oiz9aaL_xcZd-x0qd8G0VN

Репозиторий датасетов UCI: <https://archive.ics.uci.edu/>

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Личный кабинет преподавателя или студента БГУ <https://my.bsu.ru/>

GitHub (веб-сервис для хостинга IT-проектов и их совместной разработки):

<https://github.com/>

ANACONDA (дистрибутив языков программирования Python и R, включающий набор популярных свободных библиотек, объединённых проблематиками науки о данных и машинного обучения): <https://www.anaconda.com/>

Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Кабинет для лекционных занятий с мультимедийным оборудованием.

Компьютерный класс с мультимедийным оборудованием.

ПРИЛОЖЕНИЕ

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова»
Институт математики, физики и компьютерных наук
Кафедра информационных систем и методов искусственного интеллекта

Фонд оценочных средств по учебной дисциплине

Прикладные задачи машинного обучения

Направление подготовки
02.04.03 Математическое обеспечение и администрирование информационных систем

Квалификация
Магистр

Форма обучения
Очная

Улан-Удэ
2025

Паспорт фонда оценочных средств (ФОС) по дисциплине «Прикладные задачи машинного обучения»

Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

ПК-2	Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач
ПК-2.2	Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий
ПК-2.3	Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области
ПК-3	Способен руководить проектами по созданию комплексных систем искусственного интеллекта
ПК-3.1	Применяет методы и средства разработки систем искусственного интеллекта
ПК-3.2	Руководит разработкой архитектуры комплексных систем искусственного интеллекта

Этапы формирования компетенции

Семестр	Вид контроля	Оценочные средства
2 семестр	Текущий	Лабораторные работы
	Итоговый (экзамен)	Теоретические вопросы к экзамену

Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Фонд оценочных средств сформирован на основе ключевых принципов оценивания:

- валидность: объекты оценки должны соответствовать поставленным целям обучения;
- определенность: оценочные средства должны быть понятны каждому обучающемуся;
- однозначность: одинаковость оценки качества оценочного средства;
- надежность: использование единообразных показателей и критериев для оценивания достижений.

Описание показателей и критериев оценивания компетенций, а также шкал оценивания

Показатели оценивания компетенций	Уровень сформированности компетенций	Шкала оценивания	Официальный цифровой эквивалент оценки
Знать: - определения основных понятий, содержательную структуру и типизацию прикладных задач машинного обучения; - основные методологические подходы, принципы и алгоритмы построения и обучения различных моделей прикладных задач; - комплекс эвристик для	Высокий	85 – 100 баллов	5 (отлично)
	Базовый	70 – 84 баллов	4 (хорошо)
	Пороговый	60-69 баллов	3 (удовлетворительно)

<p>повышения качества моделей.</p> <p>Уметь:</p> <ul style="list-style-type: none"> - формализовывать прикладные задачи машинного обучения; - выбирать в соответствии условиям задачи модели, методы и технологии ее решения; - решать прикладные задачи машинного обучения с применением технологий программирования <p>Владеть:</p> <ul style="list-style-type: none"> - современным понятийным аппаратом машинного обучения, в т.ч. глубокого обучения; - современными инструментальными средствами и технологиями программной реализации решения прикладных задач. 			
---	--	--	--

Балльно-рейтинговая система

Для текущего и итогового контроля качества обучения студентов и магистрантов применяется балльно-рейтинговая система, разработанная в соответствии с «Положением об организации учебного процесса с применением кредитно-модульной системы обучения», утвержденным Учебно-методическим советом ФГБОУ ВО «Бурятский государственный университет» от 20 февраля 2012 г. Целью БРС является определение уровня успешности освоения (завершения изучения) обучающимися учебных дисциплин (модулей, циклов) через балльные оценки и рейтинги качества сформированных знаний, умений, профессиональных компетенций, накапливаемые в соответствии с измеряемыми в зачетных единицах трудоемкостями каждого цикла (модуля, дисциплины) и основной образовательной программы в целом.

1. Общая максимальная сумма баллов, которую студент может набрать по дисциплине в течение семестра – 100 баллов: 60 баллов текущий контроль и 40 баллов итоговый контроль (экзамен).
2. Минимальная сумма баллов, при которой студент допускается к экзамену (итоговому контролю), равна 20 баллам.
3. Минимальная сумма баллов, при которой студент получает положительную итоговую оценку по дисциплине равна 60 баллам (60% от 100 баллов).
4. Максимальная оценка за выполнение одной лабораторной работы – 10 баллов.

Связь между четырехбалльной и стобалльной системами оценки качества обучения студентов

Оценка	Рейтинговые баллы
Отлично	80-100
Хорошо	70-80
Удовлетворительно	60-70
Неудовлетворительно	<60

ПРИМЕРЫ ЛАБОРАТОРНЫХ РАБОТ

Лабораторная работа №1. Преобразование Фурье и сингулярно-спектральное разложение импульсного сигнала

Задание. Определить и сравнить частоты главных компонент Фурье-разложения с частотами основных гармоник сингулярно-спектрального разложения (SSA) на выборке импульсных сигналов.

Вход: Выборка импульсных сигналов объема N . Частота дискретизации сигнала 100 Гц. Двухбайтовая кодировка.

Выход:

- N значений частоты главной компоненты Фурье-разложения
- N значений частоты главной гармоники сингулярно-спектрального разложения (SSA)
- График их зависимости: абсцисса – частота по Фурье, ордината – частота по SSA, линейный тренд – биссектриса первой координатной четверти

Требования к реализации:

- модульная структура программы
- комментарии в коде программы

Методические рекомендации.

1. Применение стандартных пакетов Python для анализа данных (numpy, SciPy, pandas, matplotlib),

2. Вариант чтения файла с сигналом:

```
import numpy as np
from struct import unpack
F = open(filename, 'br')
signal = np.array(unpack('10000H', F.read()))
```

3. Формула прямого преобразования Фурье: $X_n = \sum_{n=0}^{N-1} x_n \left(\cos \frac{2\pi kn}{N} + i \cdot \sin \frac{2\pi kn}{N} \right)$, где N

– количество значений ряда (сигнала), x_n – значения исходного ряда, $\frac{|x_n|}{N}$ – амплитуда n -го синусоидального сигнала, $\arg(X_n)$ – фаза n -го синусоидального сигнала, k – индекс (номер) частоты, $\frac{k}{T}$ – частота k -го сигнала, T – длительность сигнала в секундах.

4. Метод SSA (Singular spectrum analysis) пошагово:

[https://ru.wikipedia.org/wiki/SSA_\(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4\)](https://ru.wikipedia.org/wiki/SSA_(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4))

5. Возможный инструментальный для реализации SSA на Python:

<https://pyts.readthedocs.io/en/stable/generated/pyts.decomposition.SingularSpectrumAnalysis.html#pyts.decomposition.SingularSpectrumAnalysis>

Литература.

1. Разделы 2.1 и 2.2 лекционного материала курса «Методы искусственного интеллекта в восточной медицине».

2. Сингулярное разложение:

http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D0%B8%D0%BD%D0%B3%D1%83%D0%BB%D1%8F%D1%80%D0%BD%D0%BE%D0%B5_%D1%80%D0%B0%D0%B7%D0%BB%D0%BE%D0%B6%D0%B5%D0%BD%D0%B8%D0%B5

3. Поучительная песенка про SVD (приятного прослушивания):

<https://www.youtube.com/watch?v=JEYLfIVvR9I>

Лабораторная работа №2. Вейвлет-преобразование импульсного сигнала

Задание. Построить вейвлет-спектры импульсных сигналов по K различным (на выбор) базисным вейвлетам.

Вход: Выборка импульсных сигналов объема N . Частота дискретизации сигнала 100 Гц. Тип файлов – двоичная двухбайтовая.

Выход: K каталогов (свой каталог у каждого базисного вейвлета), в каждом из которых N вейвлет-спектров импульсных сигналов разбитых на два подкаталога по половому признаку носителя сигнала.

Требования к реализации:

- модульная структура программы
- комментарии в коде программы

Методические рекомендации.

1. Дискретное вейвлет-преобразование (ДВП) оперирует с дискретными значениями параметров a и b , которые задаются, как правило, в виде степенных функций:

$$a = a_0^{-m}, b = k \cdot a_0^{-m}, a_0 > 1, m, k \in Z$$

где Z – пространство целых чисел $\{-\infty, \infty\}$,

m – параметр масштаба,

k – параметр сдвига.

Базис пространства $L^2(R)$ в дискретном представлении:

$$\psi_{mk} = |a_0|^{-\frac{m}{2}} \psi(a_0^{-m}t - k), m, k \in Z, \psi(t) \in L^2(R)$$

Вейвлет-коэффициенты прямого преобразования:

$$C_{mk} = \int_{-\infty}^{\infty} s(t) \psi_{mk}(t) dt$$

В общем случае, значение a может быть произвольным, но обычно принимается равным 2

2. Рекомендуемая библиотека PyWavelets: <https://pywavelets.readthedocs.io/en/latest/#main-features>

Литература.

1. Разделы 2.3 и 2.4 лекционного материала курса «Методы искусственного интеллекта в восточной медицине».
2. Вейвлет преобразование (Материал из Национальной библиотеки им. Н. Э. Баумана): <https://ru.bmstu.wiki/Вейвлет-преобразование>
3. Доступное объяснение ДВП: https://www.youtube.com/watch?v=56lr_2KqzsU
4. Примеры реализации (применения): <https://habr.com/ru/post/451278/>

ПРИМЕРЫ ДОПОЛНИТЕЛЬНЫХ ЛАБОРАТОРНЫХ РАБОТ

Цель лабораторных работ:

Получение практических навыков обработки текста и графовых структур с использованием векторных представлений, понимание принципов преобразования исходных данных в численные вектора и оценки качества моделей на реальных примерах.

Задача 1. TF-IDF преобразование текста

Цель задания: Понимание принципа весового распределения слов на основе частотности и редкости терминов.

Задача:

Выберите корпус текстов объемом минимум 5 документов (например, аннотаций статей, книг, научных публикаций). Реализуйте вычисление матрицы TF-IDF вручную и сравните её с результатом библиотеки scikit-learn. Постройте график частоты появления ключевых терминов.

Вопросы для проверки понимания:

- Почему некоторые термины имеют низкий коэффициент IDF?
- Какие преимущества даёт использование TF-IDF перед простым подсчетом встречаемости слов?

Задача 2. Word Embedding с использованием алгоритма word2vec

Цель задания: Ознакомление с концепцией встраивания слов в пространство признаков.

Задача:

Используя библиотеку `gensim`, создайте собственную реализацию модели `word2vec` на небольшом корпусе текстов. Оцените качество полученного эмбединга путём анализа ближайших соседей заданных слов и визуализации с помощью `t-SNE` или `PCA`.

Вопросы для проверки понимания:

- Какова роль контекста в обучении `word2vec`?
- Что означает понятие "близость" двух слов в пространстве эмбедингов?

Задача 3. Представление графа в виде матриц смежности и инцидентности

Цель задания: Освоение способов хранения и манипулирования графовыми структурами.

Задача:

Создайте простой направленный граф с произвольным числом вершин и рёбер.

Реализуйте методы построения матриц смежности и инцидентности. Напишите алгоритм обхода всех путей длиной N шагов из выбранной вершины A .

Вопросы для проверки понимания:

- Чем отличаются матрица смежности и матрица инцидентности?
- Для каких типов задач удобней использовать каждую из этих форматов?

Задача 4. Graph Embeddings

Цель задания: Применение методов `graph embeddings` для изучения характеристик узлов и связей.

Задача:

Реализовать метод `Node2Vec` на простом ориентированном графе. Проанализируйте распределение узлов в двумерном пространстве признаков. Объясните влияние параметра p/q на полученный результат.

Вопросы для проверки понимания:

- Что означают параметры p и q в `Node2Vec`?
- Как изменение этих параметров влияет на структуру полученных эмбедингов?

Требования к отчету:

Отчёт должен содержать подробное описание каждой части заданий, код реализации каждого шага, пояснения используемых формул и подходов, выводы по каждому пункту. Отчёт должен включать графики и таблицы результатов. Обязательно сформулируйте собственные наблюдения и рекомендации по улучшению решений.

Критерии оценивания:

- Выполнение всех пунктов лабораторных работ (80%)
- Ясность изложения материала и аккуратность оформления отчёта (10%)

- Правильность и полнота выводов (10%)

СПИСОК ВОПРОСОВ НА ЭКЗАМЕН

1. Что такое машинное обучение?
2. Чем отличается глубокое обучение от классического машинного обучения?
3. Какие типы алгоритмов машинного обучения существуют?
4. Объясните разницу между контролируемым и неконтролируемым обучением.
5. Приведите примеры задач классификации и регрессии.
6. Как классифицируются методы обработки признаков?
7. Назначение кросс-валидации и её виды.
8. Понятие переобучения и недополнения моделей.
9. Для чего применяется регуляризация в моделях машинного обучения?
10. Основные этапы построения ML-модели.
11. Зачем необходима предварительная обработка данных перед построением модели?
12. Какие способы нормализации данных существуют?
13. Опишите метод One-Hot Encoding и его применение.
14. Когда и зачем используется стандартизация данных?
15. Что такое выбросы (outliers)? Способы обнаружения и удаления выбросов.
16. Методы заполнения пропусков в данных.
17. Какие методы снижения размерности пространства признаков известны?
18. Преимущества и недостатки метода PCA.
19. Особенности методов t-SNE и UMAP для визуализации многомерных данных.
20. Что такое балансировка классов? Основные подходы.
21. Дайте определение линейной регрессии и её разновидности.
22. Принцип работы логистической регрессии.
23. Расскажите о работе деревьев решений и случайных лесов.
24. Чем отличаются бустинг и бэггинг?
25. Какой алгоритм лежит в основе градиентного бустинга?
26. Суть k ближайших соседей (kNN).
27. Почему SVM считается мощным инструментом для классификации?
28. Примеры использования ансамблевых методов.
29. В чём особенность метода кластеризации K-means?
30. Какие техники используются для оценки качества кластеризаций?
31. Назовите наиболее распространённые метрики точности для задач классификации.
32. Из каких компонентов состоит матрица ошибок (confusion matrix)?
33. Как вычисляется точность (accuracy), полнота (recall), F1-мера и специфичность (specificity)?
34. Описать понятие AUC-ROC кривой и её интерпретацию.
35. Какие показатели применяются для оценки качества регрессоров?
36. Какие ограничения имеет показатель accuracy?
37. Применение precision-recall curve и mAP-метрик.
38. Метрика Mean Average Precision (mAP). Её назначение и особенности расчёта.
39. Отличия показателей MSE, RMSE и MAE.
40. Для чего используют коэффициент детерминации (R^2)?
41. Как выбрать оптимальные гиперпараметры модели?
42. Назовите современные библиотеки Python для машинного обучения.
43. Области применения алгоритма Random Forest.
44. Где чаще всего применяют искусственные нейронные сети?
45. Назовите области применения временных рядов и рекомендательных систем.
46. Задача сегментации клиентов. Подходящие модели и метрики.
47. Использование моделей машинного обучения в медицине.

48. Как применяются нейросети в обработке изображений?
49. Примеры промышленных применений компьютерного зрения.
50. Машинное обучение в финансовой сфере.

Критерии оценки ответа на теоретический вопрос

Критерий	Требования	Баллы
Знание и понимание теоретического материала	Рассматриваемые понятия определяются четко и полно, приводятся соответствующие примеры	10
	Используемые понятия строго соответствуют теме	8
Анализ и оценка информации	Объясняются альтернативные взгляды на рассматриваемую проблему	8
	Дается личная оценка проблеме	6
Построение суждений	Изложение ясное и четкое	4
	Приводятся различные точки зрения и их личная оценка (примеры применения, плюсы и минусы)	4
	Итого	40