Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

Кафедра прикладной математики

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б1.Д.В.4 Машинное обучение»

Уровень высшего образования

МАГИСТРАТУРА

Направление подготовки
09.04.02 Информационные системы и технологии
(код и наименование направления подготовки)

Искусственный интеллект в промышленности
(наименование направленности (профиля) образовательной программы)

Квалификация
Магистр

Форма обучения
Очная

Год набора 2025

Рабочая программа дисциплины «*Б1.Д.В.4 Машинное обучение*» рассмотрена и утвержде
на заседании кафедры

Кафедра прикладной математики
наименование кафедры

протокол № 6 от "30" января 2025 г.

Заведующий кафедрой
Кафедра прикладной математики _____ И.П. Болодурина
наименование кафедры *подпись* *расшифровка подписи*

Исполнители:
Доцент _____ С.Т. Дусакаева
должность *подпись* *расшифровка подписи*

_____
должность *подпись* *расшифровка подписи*

СОГЛАСОВАНО:
Председатель методической комиссии по направлению подготовки
09.04.02 Информационные системы и технологии _____ И.П. Болодурина
код наименование *личная подпись* *расшифровка подписи*

Научный руководитель магистерской программы _____ И.П. Болодурина
личная подпись *расшифровка подписи*

Заведующий отделом формирования фонда и научной обработки документов
_____ Н.Н. Бигалиева С.А. Бикмиллеева
личная подпись *расшифровка подписи*

Уполномоченный по качеству факультета
Доцент _____ С.Н. Морозова
личная подпись *расшифровка подписи*

№ регистрации _____

© Дусакаева С.Т., 2025
© ОГУ, 2025

1 Цели и задачи освоения дисциплины

Цель (цели) освоения дисциплины: сформировать у студентов представление о наборе технологий и методов интеллектуального анализа (Data Mining), предназначенных для обработки исходных данных за счет автоматизации процесса извлечения новых, корректных и потенциально полезных знаний.

Задачи:

- формирование научного представления о современных методах интеллектуального анализа данных;
- умения анализировать задачи машинного обучения и осуществлять взвешенный выбор того или иного решения;
- приобретение знаний о принципах и алгоритмах, лежащих в основе современных интеллектуальных систем анализа данных;
- овладение технологиями и методиками сбора, предварительной подготовки и анализа экспериментальных данных;
- приобретение практических навыков работы с конкретными программными средствами интеллектуального анализа данных;
- использование изученных языков программирования в качестве основы для разработки различных алгоритмов машинного обучения.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к обязательным дисциплинам (модулям) вариативной части блока Д «Дисциплины (модули)»

Пререквизиты дисциплины: *Б1.Д.Б.6 Теоретическая информатика, Б1.Д.Б.7 Объектно-ориентированные технологии и базы данных*

Постреквизиты дисциплины: *Б1.Д.В.6 Глубокое обучение, Б1.Д.В.7 Автоматическое машинное обучение, Б1.Д.В.8 Интеллектуальные технологии в промышленности полного цикла, Б2.П.В.П.1 Преддипломная практика*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
ПК*-5 Способен разрабатывать оригинальные алгоритмы и программные средства с использованием современных интеллектуальных компьютерных технологий, для решения задач в области создания и применения искусственного интеллекта	ПК*-5-В-1 Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта	**Знать:** о методах, приемах, алгоритмах и способах сбора, обработки и интерпретации данных современных научных исследований в экономике. **Уметь:** собирать и обрабатывать статический,

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
		экспериментальный, теоретический, графический и т.п. материал. **Владеть:** Навыками применения современных информационно-коммуникационных и интеллектуальных технологий для решения задач в области создания и применения искусственного интеллекта.
ПК*-6 Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач искусственного интеллекта	ПК*-6-В-1 Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области ПК*-6-В-2 Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области ПК*-6-В-3 Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий	**Знать:** о современных математических методах и системах интеллектуального анализа данных. **Уметь:** Применять современные математические методы и системы интеллектуального анализа данных при решении задач в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности. **Владеть:** навыками поиска, выработки и применения новых решений на основе информационной и библиографической культуры с использованием инструментов интеллектуального анализа данных в области информационно-коммуникационных технологий.
ПК*-13 Способен понимать фундаментальные принципы работы современных систем искусственного интеллекта,	ПК*-13-В-1 Применяет современные методы и инструменты для представления результатов научно-исследовательской деятельности	**Знать:** современные методы и инструменты для представления результатов научно-

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
разрабатывать правила и стандарты взаимодействия человека и искусственного интеллекта и использовать их в социальной и профессиональной деятельности		исследовательской деятельности. **Уметь:** применять современные методы и инструменты для представления результатов научно-исследовательской деятельности. **Владеть:** навыками работы с научными источниками информации.

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единиц (180 академических часов).

Вид работы	Трудоемкость, академических часов	
	2 семестр	всего
Общая трудоёмкость	**180**	**180**
Контактная работа:	**36,5**	**36,5**
Лекции (Л)	18	18
Лабораторные работы (ЛР)	16	16
Консультации	1	1
Индивидуальная работа и инновационные формы учебных занятий	1	1
Промежуточная аттестация (зачет, экзамен)	0,5	0,5
Самостоятельная работа: *- выполнение курсовой работы (КР);* *- выполнение индивидуального творческого задания (ИТЗ);* *- самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий;* *- изучение разделов курса в системе электронного обучения;* *- подготовка к лабораторным занятиям;* *- подготовка к рубежному контролю и т.п.)*	**143,5** +	**143,5**
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	**экзамен**	

Разделы дисциплины, изучаемые в 2 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Введение в интеллектуальный анализ данных	22	2		2	18
2	Линейный классификатор и стохастический	22	2		2	18

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
	градиент					
3	Нейронные сети: градиентные методы оптимизации	24	4		2	18
4	Метрические методы классификации и регрессии	20	2		2	16
5	Метод опорных векторов	20	2		2	16
6	Многомерная линейная регрессия	22	2		2	18
7	Нелинейная регрессия	22	2		2	18
8	Критерии выбора моделей и методы отбора признаков	28	2		2	24
	Итого:	180	18		16	146
	Всего:	180	18		16	146

4.2 Содержание разделов дисциплины

1 Введение в интеллектуальный анализ данных

Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные.

Типы задач: классификация, регрессия, прогнозирование, ранжирование.

Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.

Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.

Примеры прикладных задач. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных.

2 Линейный классификатор и стохастический градиент

Линейный классификатор, модель МакКаллока-Питтса, непрерывные аппроксимации пороговой функции потерь.

Метод стохастического градиента SG.

Метод стохастического среднего градиента SAG.

Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, «выбивание» из локальных минимумов.

Проблема мультиколлинеарности и переобучения, регуляризация или редукция весов (weight decay).

Вероятностная постановка задачи классификации. Принцип максимума правдоподобия.

Вероятностная интерпретация регуляризации, совместное правдоподобие данных и модели. Принцип максимума апостериорной вероятности.

Гауссовский и лапласовский регуляризаторы.

Логистическая регрессия. Принцип максимума правдоподобия и логарифмическая функция потерь. Метод стохастического градиента для логарифмической функции потерь. Многоклассовая логистическая регрессия. Регуляризованная логистическая регрессия. Калибровка Платта.

3 Нейронные сети: градиентные методы оптимизации

Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации.

Проблема полноты. Задача исключающего или. Полнота двухслойных сетей в пространстве булевых функций.

Алгоритм обратного распространения ошибок.

Быстрые методы стохастического градиента: Поляка, Нестерова, AdaGrad, RMSProp, AdaDelta, Adam, Nadam, диагональный метод Левенберга-Марквардта.

Проблема взрыва градиента и эвристика gradient clipping.

Метод случайных отключений нейронов (Dropout). Интерпретации Dropout. Обратный Dropout и L2-регуляризация.

Функции активации ReLU и PReLU. Проблема «паралича» сети.

Эвристики для формирования начального приближения. Метод послойной настройки сети.

Подбор структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage).

4 Метрические методы классификации и регрессии

Гипотезы компактности и непрерывности.

Обобщённый метрический классификатор.

Метод ближайших соседей kNN и его обобщения. Подбор числа k по критерию скользящего контроля.

Метод окна Парзена с постоянной и переменной шириной окна.

Метод потенциальных функций и его связь с линейной моделью классификации.

Задача отбора эталонов. Полный скользящий контроль (CCV), формула быстрого вычисления для метода 1NN. Профиль компактности.

Отбор эталонных объектов на основе минимизации функционала полного скользящего контроля.

Непараметрическая регрессия. Локально взвешенный метод наименьших квадратов. Ядерное сглаживание.

Оценка Надарая-Ватсона с постоянной и переменной шириной окна. Выбор функции ядра и ширины окна сглаживания.

Задача отсева выбросов. Робастная непараметрическая регрессия. Алгоритм LOWESS.

5 Метод опорных векторов

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).

Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь.

Задача квадратичного программирования и двойственная задача. Понятие опорных векторов.

Рекомендации по выбору константы C.

Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера.

Способы конструктивного построения ядер. Примеры ядер.

SVM-регрессия.

Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.

Метод релевантных векторов RVM.

6 Многомерная линейная регрессия

Задача регрессии, многомерная линейная регрессия.

Метод наименьших квадратов, его вероятностный смысл и геометрический смысл.

Сингулярное разложение.

Проблемы мультиколлинеарности и переобучения.

Регуляризация. Гребневая регрессия через сингулярное разложение.

Методы отбора признаков: Лассо Тибширани, Elastic Net, сравнение с гребневой регрессией.

Метод главных компонент и декоррелирующее преобразование Карунена-Лоэва, его связь с сингулярным разложением.

Спектральный подход к решению задачи наименьших квадратов.

Задачи и методы низкоранговых матричных разложений.

7 Нелинейная регрессия

Метод Ньютона-Рафсона, метод Ньютона-Гаусса.

Обобщённая аддитивная модель (GAM): метод настройки с возвращениями (backfitting) Хасти-Тибширани.

Логистическая регрессия. Метод наименьших квадратов с итеративным пересчётом весов (IRLS). Пример прикладной задачи: кредитный скоринг. Бинаризация признаков. Скоринговые карты и оценивание вероятности дефолта. Риск кредитного портфеля банка.

Обобщённая линейная модель (GLM). Экспоненциальное семейство распределений.

Неквадратичные функции потерь. Метод наименьших модулей. Квантильная регрессия. Пример прикладной задачи: прогнозирование потребительского спроса.

Робастная регрессия, функции потерь с горизонтальными асимптотами.

8 Критерии выбора моделей и методы отбора признаков

Критерии качества классификации: чувствительность и специфичность, ROC-кривая и AUC, точность и полнота, AUC-PR.

Внутренние и внешние критерии. Эмпирические и аналитические критерии.

Скользящий контроль, разновидности эмпирических оценок скользящего контроля. Критерий непротиворечивости.

Разновидности аналитических оценок. Регуляризация. Критерий Акаике (AIC). Байесовский информационный критерий (BIC). Оценка Вапника-Червоненкиса.

Сложность задачи отбора признаков. Полный перебор.

Метод добавления и удаления, шаговая регрессия.

Поиск в глубину, метод ветвей и границ.

Усечённый поиск в ширину, многорядный итерационный алгоритм МГУА.

Генетический алгоритм, его сходство с МГУА.

Случайный поиск и Случайный поиск с адаптацией (СПА).

4.3 Лабораторные работы

№ ЛР	№ раздела	Наименование лабораторных работ	Кол-во часов
1	1	Анализ данных на примере набора Ирисов Фишера.	2
2	2	Анализ метода стохастического градиента.	2
3	3	Нейронные сети: Autograd - обучение персептрона на выборке MNIST	2
4	4	Метрические методы классификации и регрессии: метод ближайших соседей, метод Парзеновского окна и метод потенциальных функций (реализация)	2
5	5	Линейные методы классификации и регрессии: метод опорных векторов	2
6	6	Многомерная линейная регрессия. Метод главных компонент	2
7	7	Нелинейная регрессия. Обобщённые линейные модели. Нестандартные функции потерь.	2
8	8	Критерии выбора моделей и методы отбора признаков.	2
		Итого:	16

4.4 Курсовая работа (2 семестр)

Примерные темы курсовой работы:

1. Методы построения ансамблей моделей для прогнозирования цен на недвижимость
2. Методы и инструменты предиктивной аналитики при интеллектуальном анализе данных о поведении клиентов банка
3. Интеллектуальные методы анализа тональности текста в социальных сетях
4. Исследование интеллектуальных алгоритмов сегментации, выделения признаков и классификации изображений популяционной группы
5. Методы и алгоритмы предсказательного моделирования карьерного роста на основе анализа вакансий
6. Предиктивное моделирование цен на сельскохозяйственное сырье

7. Исследование эффективности методов классификации при анализе данных о заболеваемости COVID-19

8. Сравнительный анализ эффективности методов классификации на несбалансированных наборах данных по сердечно-сосудистым заболеваниям

9. Методы и алгоритмы предсказательного моделирования при исследовании цен на автомобили

10. Исследование данных о трафике в аэропортах в условиях пандемии методами искусственного интеллекта.

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Чубукова, И.А. Data Mining / И.А. Чубукова. - 2-е изд., испр. - Москва : Интернет-Университет Информационных Технологий, 2008. - 383 с. - Режим доступа: http://biblioclub.ru/index.php?page=book&id=233055

2. Интеллектуальные системы: учебное пособие/ А. М. Семенов, Н. А. Соловьев, Е. Н. Черно-прудова, А. С. Цыганков; М-во образования и науки Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. проф. образования "Оренбург. гос. ун-т". - Оренбург: ОГИМ, 2014. - 237 с.

5.2 Дополнительная литература

1. Крутиков, В.Н. Анализ данных: учебное пособие / В.Н. Крутиков, В.В. Мешечкин - Кемерово: Кемеровский государственный университет, 2014. – 138 с. Режим доступа: http://biblioclub.ru/index.php?page=book&id=278426

2. Павлов, С. И. Системы искусственного интеллекта. Часть 1. Учебное пособие [Электронный ресурс] / Павлов С. И. - Томский государственный университет систем управления и радиоэлектроники, 2011. Режим доступа: http://biblioclub.ru/index.php?page=book&id=208933

5.3 Периодические издания

1. Вестник компьютерных и информационных технологий
2. Информационные технологии

5.4 Интернет-ресурсы

1. http://machinelearning.ru - Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.

2. https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie/ «Coursera»; МООК: «Введение в машинное обучение»

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. Операционная система RedOs.
2. Интегрированная система решения математических, инженерно-технических и научных задач PTC MathCAD 14.0.
3. Автоматизированная интерактивная система сетевого тестирования - АИССТ (зарегистрирована в РОСПАТЕНТ, Свидетельство о государственной регистрации программы для ЭВМ №2011610456, правообладатель – Оренбургский государственный университет), режим доступа - http://aist.osu.ru.
4. 7-Zip. Предоставляется по лицензии GNU LGPL. Разработчик: Игорь Павлов. Режим доступа: http://www.7-zip.org/.
5. Adobe Acrobat Reader DC. Доступно бесплатно после принятия условий лицензионного соглашения на ПО Adobe. Разработчик: Adobe Systems. Режим доступа: https://get.adobe.com/ru/reader/.
6. ГАРАНТ Платформа F1 [Электронный ресурс]: справочно-правовая система. / Разработчик ООО НПП «ГАРАНТ-Сервис», 119992, Москва, Воробьевы горы, МГУ, [1990–2016]. – Режим доступа в сети ОГУ для установки системы: \\fileserver1\GarantClient\garant.exe.

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, семинарского типа, курсового проектирования, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения лабораторных занятий используется лаборатория «*Наименование*» (при наличии), (компьютерный класс) оснащенная/ оснащенный (указывается конкретное оборудование и т.п.)

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.