

Методы машинного обучения

рабочая программа дисциплины (модуля)

Закреплена за кафедрой **Экономики минерально-сырьевого комплекса**

Учебный план b010305_25_BA25.plx
Направление подготовки 01.03.05 Статистика

Квалификация **Бакалавр**

Форма обучения **очная**

Общая трудоемкость **4 ЗЕТ**

Часов по учебному плану 144
в том числе:
аудиторные занятия 86,35
самостоятельная работа 30,65
часов на контроль 27

Виды контроля в семестрах:
экзамены 8

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	8 (4.2)		Итого	
Неделя	14 5/6			
Вид занятий	УП	РП	УП	РП
Лекции	28	28	28	28
Практические	56	56	56	56
Иные виды контактной работы	2,35	2,35	2,35	2,35
В том числе инт.	8	8	8	8
Итого ауд.	86,35	86,35	86,35	86,35
Контактная работа	86,35	86,35	86,35	86,35
Сам. работа	30,65	30,65	30,65	30,65
Часы на контроль	27	27	27	27
Итого	144	144	144	144

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1	Целью освоения дисциплины "Методы машинного обучения" является изучение математических основ теории и методов машинного обучения (machine learning) в вероятностной постановке.
-----	---

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Цикл (раздел) ОП:		Б1.В
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Теория статистики	
2.1.2	Микроэкономика	
2.1.3	Макроэкономика	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Основы внешнеэкономической деятельности	
2.2.2	Государственное регулирование экономики	
2.2.3	Статистика финансов	
2.2.4	Управление проектами	
2.2.5	Основы бухгалтерского учета и финансовой отчетности	
2.2.6	Микроэкономическая статистика	
2.2.7	Экономика данных	
2.2.8	Системный анализ	

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ПК-1: Способен собирать данные в ходе непосредственного наблюдения и измерения с применением технических средств в соответствии с утвержденными программами, правилами и методиками

Знать:

Уровень 1	Актуальные научные публикации по статистике, в том числе зарубежные
Уровень 2	Методологические подходы к проведению экспериментальных расчетов
Уровень 3	Статистические пакеты прикладных программ

Уметь:

Уровень 1	Работать с различными источниками статистической информации
Уровень 2	Производить статистические расчеты на основе соответствующих математических и технических средств
Уровень 3	Применять статистические пакеты прикладных программ

Владеть:

Уровень 1	Навыками разработки и совершенствования методологии сбора и обработки статистических данных
Уровень 2	Навыками разработки и совершенствования систем статистических показателей и методик их расчета
Уровень 3	Навыками самостоятельного ведения научного поиска и публикации результатов научной деятельности в области математической статистики

В результате освоения дисциплины (модуля) обучающийся должен

3.1	Знать:
3.1.1	о методах, приемах, алгоритмах и способах сбора, обработки и интерпретации данных;
3.1.2	основные методы и алгоритмы машинного обучения для решения поставленных задач;
3.1.3	возможности инструментальных средств для анализа данных;
3.1.4	основные типы задач машинного обучения для последующего выбора методов обработки информации
3.2	Уметь:
3.2.1	собирать и обрабатывать статистический, экспериментальный, теоретический, графический и т.п. материал;
3.2.2	применять основные модели машинного обучения для анализа и синтеза информации с использованием компьютерных технологий;
3.2.3	применять технические средства для оценки параметров различных процессов и явлений, являющихся объектами статистических наблюдений;
3.2.4	разрабатывать прикладные программы анализа данных;
3.2.5	формализовать знания экспертов с применением различных методов представления знаний
3.3	Владеть:
3.3.1	проведением качественного и количественного анализа собранных данных;

3.3.2	методами разработки алгоритмов и программного обеспечения работы основных типов;
3.3.3	инструментальными средствами анализа данных;
3.3.4	техникой предварительной обработки данных, в том числе перенесения информации на электронные носители и по каналам связи, для проведения последующего анализа данных

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература	Инте ракт.	Примечание
	Раздел 1. Тема 1. Основные понятия и примеры прикладных задач						
1.1	Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Вероятностная постановка задачи обучения. Принцип максимума правдоподобия. Проблема переобучения и понятие обобщающей способности. Типы задач: классификация, регрессия, ранжирование, прогнозирование. /Лек/	8	4	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	4	
1.2	Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. /Пр/	8	9	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	4	
1.3	Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия. Примеры прикладных задач. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. /Ср/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
	Раздел 2. Итоговая аттестация						
2.1	Проведение консультации перед экзаменом /ИВКР/	8	2	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
2.2	Прием экзамена /ИВКР/	8	0,35	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
	Раздел 3. Тема 2. Линейные методы классификации и регрессии						
3.1	Линейный классификатор, модель МакКаллока-Питтса, непрерывные аппроксимации пороговой функции потерь. Метод стохастического градиента SG. /Лек/	8	4	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	

3.2	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. /Пр/	8	9	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
3.3	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. /Ср/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
	Раздел 4. Тема 3. Метрические методы классификации и регрессии						
4.1	Гипотезы компактности и непрерывности. Обобщенный 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. /Лек/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
4.2	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. /Пр/	8	9	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
4.3	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. /Ср/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	

	Раздел 5. Тема 4. Метод опорных векторов						
5.1	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. /Лек/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
5.2	Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера. Способы конструктивного построения ядер. Примеры ядер. SVM-регрессия. /Пр/	8	9	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
5.3	Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM. Метод релевантных векторов RVM. /Ср/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
	Раздел 6. Тема 5. Многомерная линейная регрессия						
6.1	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. /Лек/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
6.2	Методы отбора признаков: Лассо Тибширани, Elastic Net, сравнение с гребневой регрессией. Метод главных компонент и декоррелирующее преобразование Карунена-Лоэва, его связь с сингулярным разложением /Пр/	8	10	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
6.3	Спектральный подход к решению задачи наименьших квадратов. Задачи и методы низкоранговых матричных разложений. /Ср/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
	Раздел 7. Тема 6. Критерии выбора моделей и методы отбора признаков						
7.1	Критерии качества классификации: чувствительность и специфичность, ROC-кривая и AUC, точность и полнота, AUC-PR. Внутренние и внешние критерии. Эмпирические и аналитические критерии. Скользкий контроль, разновидности эмпирических оценок скользящего контроля. Критерий непротиворечивости /Лек/	8	5	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	

7.2	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. /Пр/	8	10	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	
7.3	Усечённый поиск в ширину, многорядный итерационный алгоритм МГУА. Генетический алгоритм, его сходство с МГУА. Случайный поиск и Случайный поиск с адаптацией (СПА). /Ср/	8	5,65	ПК-1	Л1.1 Л1.2Л2.1 Л2.2Л3.1 Л3.2 Э1 Э2 Э3 Э4	0	

5. ОЦЕНОЧНЫЕ СРЕДСТВА

5.1. Контрольные вопросы и задания

Контрольные вопросы для подготовки к промежуточной аттестации:

1. Формальная модель машинного обучения.
2. Функционал качества в обучении по прецедентам.
3. Модели обучения по прецедентам.
4. Обобщенный метрический классификатор: принцип работы и примеры применения.
5. Метод ближайшего соседа в машинном обучении.
6. Метод потенциальных функций.
7. Метод парзеновского окна.
8. Оптимальный байесовский классификатор.
9. Наивный байесовский классификатор.
10. Нормальный дискриминантный анализ.
11. ЕМ - алгоритм
12. Линейный классификатор и линейная регрессия.
13. Метод стохастического линейного спуска.
14. Логистическая регрессия.
15. Машина опорных векторов.
16. Решающие деревья в машинном обучении.
17. Нейронные сети в машинном обучении.
18. Сверточные сети в задачах машинного обучения.

Задания для текущего контроля представлены в Приложении 1.

5.2. Темы письменных работ

К письменным работам по дисциплине «Методы машинного обучения» относятся рефераты.

Примерные темы рефератов:

1. Алгоритмы фильтрации «зашумленных» данных
2. Алгоритмы предварительной обработки наборов данных для машинного обучения
3. Современные прикладные пакеты машинного обучения
4. Платформы машинного обучения (от «Сбер», «Яндекс» и др.)
5. Варианты применения машинного обучения в образовательных процессах
6. Новые алгоритмы деревьев решений
7. Варианты применения бустинга и бэггинга в практических задачах

5.3. Оценочные средства

Рабочая программа учебной дисциплины «Методы машинного обучения» обеспечена оценочными средствами для проведения текущего контроля и промежуточной аттестации, включающими контрольные вопросы для проведения промежуточной аттестации, критерии оценивания учебной деятельности обучающихся по балльно-рейтинговой системе, примеры заданий для занятий семинарского типа, билеты для проведения промежуточной аттестации.

Все оценочные средства представлены в Приложении 1.

5.4. Перечень видов оценочных средств

Оценочные средства разработаны для всех видов учебной деятельности обучающегося – занятий семинарского типа (практических занятий), самостоятельной работы и промежуточной аттестации.

Оценочные средства представлены в виде:

- средств текущего контроля: проверки решений тестовых заданий, собеседования по теме;
- средств итогового контроля – промежуточной аттестации: экзамена в 8 семестре.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)			
6.1. Рекомендуемая литература			
6.1.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Алибеков И. Ю.	Теория вероятностей и математическая статистика в среде MATLAB: учебное пособие для вузов	Санкт-Петербург: Лань, 2024
Л1.2	Платонов А. В.	Машинное обучение: учебное пособие для вузов	Москва: Юрайт, 2024
6.1.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Иванов Б. Н.	Теория вероятностей и математическая статистика: учебное пособие для вузов	Санкт-Петербург: Лань, 2024
Л2.2	Дудин М. Н., Лясников Н. В., Лезина М. Л.	Статистика: учебник и практикум для вузов	Москва: Юрайт, 2024
6.1.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	Елисеева И. И., Боченина М. В., Бурова Н. В., Михайлов Б. А.	Статистика: учебник и практикум для вузов	Москва: Юрайт, 2024
Л3.2	Елисеева И. И., Батырова Д. К., Боченина М. В., Декина М. П., Долотовская О. В., Капралова А. В., Капралова Е. Б., Курышева С. В., Михайлов Б. А., Нерадовская Ю. В., Парик И. Ю., Потахова Л. М., Силаева С. А., Флуд Н. А.	Статистика: учебник для вузов	Москва: Юрайт, 2024
6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"			
Э1	Электронные ресурсы библиотека МГРИ		
Э2	ООО «Книжный Дом Университета» (БиблиоТех)		
Э3	ООО ЭБС Лань		
Э4	Библиографическая и реферативная база данных SCOPUS		
6.3.1 Перечень программного обеспечения			
6.3.1.1	Office Professional Plus 2019		
6.3.1.2	Windows 10		
6.3.1.3	Windows 7		
6.3.1.4	Windows 8		
6.3.1.5	ПО "Интерне-расширение информационной системы"	Автоматизация управления учебным процессом. Интернет-расширение представляет собой динамический сайт, подключаемый к единой базе данных ИС «Деканат», «Электронные ведомости», «Планы». Данная подсистема обеспечивает:	
6.3.1.6	ПО ""Визуальная студия тестирования"	Автоматизация управления учебным процессом. Позволяет автоматизировать контроль знаний студентов, включая создание набора тестовых заданий, проведение тестирования студентов и анализ результатов.	
6.3.1.7	ПО "Электронные ведомости"	Автоматизация управления учебным процессом. Предназначена для учета и анализа успеваемости студентов.	
6.3.1.8	ПО "Планы"	Автоматизация управления учебным процессом. Позволяет создать в рамках высшего учебного заведения единую систему автоматизированного планирования учебного процесса.	
6.3.1.9	ПО "Деканат"	Автоматизация управления учебным процессом. Предназначена для ведения личных дел студентов.	

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)			
Аудитория	Назначение	Оснащение	Вид
1	Специализированная многофункциональная учебная аудитория № 1 для проведения учебных занятий лекционного и семинарского типов, групповых и индивидуальных консультаций, текущего контроля и промежуточной/ итоговой аттестации	Столы обучающихся; Стулья обучающихся; Письменный стол педагогического работника; Стул педагогического работника; Кафедра; Магнитно-маркерная доска; Мультимедийный проектор; Экран; Ноутбук с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде	
3	Специализированная многофункциональная учебная аудитория № 3 для проведения учебных занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной/ итоговой аттестации	Компьютерные столы обучающихся; Стулья обучающихся; Письменный стол педагогического работника; Стул педагогического работника; Стеллаж для учебно-методических материалов, в том числе учебно-наглядных пособий; Многофункциональное устройство (принтер, сканер, ксерокс); Интерактивная доска; Мультимедийный проектор; Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде	
5	Помещение № 5 для самостоятельной работы обучающихся	Письменный стол обучающегося; Стул обучающегося; Письменный стол обучающегося с ограниченными возможностями здоровья; Стул обучающегося с ограниченными возможностями здоровья; Ноутбук с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата; Моноблок (в том числе, клавиатура, мышь, наушники) с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде	

7	Помещение № 7 для самостоятельной работы обучающихся	Столы; Стулья; Стеллаж для учебно-методических материалов; Многофункциональное устройство (принтер, сканер, ксерокс); Ноутбук с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата; Моноблок (в том числе, клавиатура, мышь, наушники) с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде	
---	--	---	--

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Методические указания по изучению учебной дисциплины «Методы машинного обучения» представлены в Приложении 2 и включают в себя:

1. Методические указания для обучающихся по организации учебной деятельности.
2. Методические указания по организации самостоятельной работы обучающихся.
3. Методические указания по организации процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.