

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

Московский государственный университет имени М.В. Ломоносова
филиал МГУ в г. Севастополе
факультет экономики и управления
кафедра экономики



УТВЕРЖДАЮ

Директор

Филиала МГУ в г. Севастополе

О.А. Шпырко

август 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

В-ПД МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ И МАШИННОЕ ОБУЧЕНИЕ

код и наименование дисциплины (модуля)

Уровень высшего образования:

бакалавриат

Направление подготовки:

38.03.01 «Экономика»

(код и название направления/специальности)

Направленность (профиль) ОПОП:

общий

(если дисциплина (модуль) относится к вариативной части программы)

Форма обучения:

очная

очная, очно-заочная

Рабочая программа рассмотрена
на заседании кафедры экономики
протокол №1 от «26» августа 2024 г.

Заведующий кафедрой

(подпись)

(Н.А. Розинская)

Рабочая программа одобрена

Методическим советом

Филиала МГУ в г. Севастополе

Протокол №10 от «29» августа 2024 г.

(Л.И. Теплова)

(подпись)

Севастополь, 2024

Рабочая программа дисциплины (модуля) разработана в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 38.03.01 Экономика (уровень бакалавриата), утвержденного Приказом Министерства науки и высшего образования Российской Федерации от 12.08.2020 № 954.

Год (годы) приема на обучение 2021

курс – 4

семестр – 7

зачетных единиц – 4

академических часов – 144, в т.ч.

лекций – 36 часа

практических занятий – 36 часа

Форма промежуточной аттестации:

экзамен в 7-м семестре

ОГЛАВЛЕНИЕ

1. Место дисциплины в структуре ОПОП ВО	4
2. Входные требования для освоения дисциплины (модуля), предварительные условия	4
3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников	4
4. Формат обучения	5
5. Объем дисциплины (модуля)	6
6. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий	6
7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю)	9
8. Ресурсное обеспечение	12
9. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП	13
10. Язык преподавания	13
11. Преподаватель (преподаватели)	13
12. Автор (авторы) программы	13

1. Место дисциплины (модуля) в структуре ОПОП ВО.

Учебная дисциплина «Многомерный статистический анализ и машинное обучение» принадлежит к базовой части профессионального цикла.

Данный курс изучается в 7-ом семестре, изучаемый курс является важным для научно-исследовательской работы бакалавра, в том числе для работы в рамках научно-исследовательского семинара.

Целью изучения дисциплины «Многомерный статистический анализ и машинное обучение» является формирование у студентов углубленных знаний в области математических методов анализа экономики; применения современных методов анализа данных для получения прогнозов; применение современных методов анализа данных в прикладных задачах

Задачи освоения дисциплины:

- Познакомить с основными алгоритмами многомерного статистического анализа;
- Познакомить с основными алгоритмами машинного обучения;
- Показать способы обработки текстов
- Обсудить примеры применения этих алгоритмов как в экономических исследованиях, так и в практических задачах

2. Входные требования для освоения дисциплины (модуля), предварительные

условия. Для успешного освоения дисциплины студенты должны изучить курсы теории вероятности, математической статистики, линейной алгебры, математического анализа, эконометрики.

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников.

Общекультурные компетенции выпускника:

ОК-7. Способность к самоорганизации и самообразованию;

Общепрофессиональные компетенции выпускника:

ОПК-2. Способность осуществлять сбор, анализ и обработку данных, необходимых для решения профессиональных задач;

ОПК-3. Способность выбрать инструментальные средства для обработки экономических данных в соответствии с поставленной задачей, проанализировать результаты расчетов и обосновать полученные выводы;

Профессиональные компетенции выпускника:

ПК-1. Способность собрать и проанализировать исходные данные, необходимые для расчета экономических и социально-экономических показателей, характеризующих деятельность хозяйствующих субъектов;

ПК-4. Способность на основе описания экономических процессов и явлений строить стандартные теоретические и эконометрические модели, анализировать и содержательно интерпретировать полученные результаты;

ПК-6. Способность анализировать и интерпретировать данные отечественной и зарубежной статистики о социально-экономических процессах и явлениях, выявлять тенденции изменения социально-экономических показателей;

ПК-7. Способность, используя отечественные и зарубежные источники информации, собрать необходимые данные, проанализировать их и подготовить информационный обзор и/или аналитический отчет;

ПК-12. Способность использовать в преподавании экономических дисциплин в образовательных учреждениях различного уровня, существующие программы и учебно-методические материалы.

Планируемые результаты обучения по дисциплине (модулю):

Знать:

- ☐ основные способы анализа многомерных статистических данных, их применение к исследованию экономических процессов
- ☐ основные задачи, решаемые стандартными алгоритмами обработки данных
- ☐ основы программирования в языке R

Уметь:

- ☐ реализация основных алгоритмов анализа данных в языке программирования R
- ☐ содержательно интерпретировать полученные результаты

Владеть:

- ☐ основы программирования в языке R

Иметь опыт:

- ☐ -самостоятельного решения практических задач прикладного анализа данных

4. Формат обучения

Преподавание дисциплины может быть реализовано в смешанном формате, очно в аудиториях учебного корпуса и с помощью программного обеспечения с закрытым кодом,

обеспечивающее текстовую, голосовую и видеосвязь через Интернет между компьютерами (Skype) <https://join.skype.com/oFU2liho0LEH>.

5. Объем дисциплины (модуля)

составляет 5 з.е., в том числе 72 академических часов, отведенных на контактную работу обучающихся с преподавателем (аудиторная нагрузка), 108 академических часов на самостоятельную работу обучающихся.

6. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий

Наименование и краткое содержание разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе			
		Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, часы			Самостоятельная работа обучающегося, часы (виды самостоятельной работы – эссе, реферат, контрольная работа и пр. – указываются при необходимости)
		Занятия лекционного о типа*	Занятия семинарског о типа*	Всего	
Тема 1. Корреляционный анализ	8	2	2	4	8
Тема 2. Классификация многомерных наблюдений без обучения	16	4	4	8	8, обязательное домашнее задание 1
Тема 3. Снижение размерности.	16	4	4	8	8
Тема 4. Дискриминантны	8	2	2	4	10, обязательное домашнее задание 1

й анализ					
Тема 5. Оптимизация	8	2	2	4	10
Тема 6. Метрические методы классификации	16	4	4	8	10
Тема 6. Решающие деревья	8	2	2	4	10
Тема 7. Метрики качества	16	4	4	8	10
Тема 8. Линейные методы классификации	16	4	4	8	10
Тема 9. Анамбли алгоритмов	16	4	4	8	10, обязательное домашнее задание 1
Тема 10. Случайный лес.	8	2	2	4	7
Тема 11. Нейросети	16	4	4	8	7, проект
Промежуточная аттестация Экзамен - защита проекта					8
Итого	<i>180</i>			<i>72</i>	108

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю)

7.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.

Домашнее задание 1

В файле - различные
характеристики банков.

Требуется сделать
классификацию

- с помощью иерархической процедуры

- с помощью метода k-средних

В первом случае в отчете поместить дерево, выбранное вами разумное разбиение на кластеры, указать выбор расстояния

Во втором случае указать полученное разбиение на группы, результаты дисперсионного анализа (ипрокомментировать их)

Домашнее задание 2.

В файле - различные характеристики авиакомпаний Требуется сделать классификацию

- с помощью иерархической процедуры

- с помощью метода k-средних

По сделанной классификации осуществить дискриминантный анализ.

Привести критерии качества разбиения. Классифицировать 1-ую авиакомпанию.

Домашнее задание 3.

Студентам выдается файл с характеристиками стран. Требуется выделить главные компоненты, интерпретировать их, произвести факторный анализ.

Контрольная работа

Студентам выдается файл с данными, необходимо произвести корреляционный, кластерный и компонентный анализ, интерпретировать результаты.

7.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

Экзамен состоит в защите проекта. Примеры темы для проектов (самостоятельных эконометрических исследований)

Вопросы к экзамену

1. Постановка задачи машинного обучения с учителем. Схема работы наивного байесовского классификатора.
2. Линейная регрессия, логистическая регрессия: когда есть единственный глобальный минимум функции потерь? Определение и свойства функции softmax.
3. Граф вычислений полносвязной двухслойной нейросети для классификации. Вычисление градиента в нем. Простой стохастический градиентный спуск (vanilla SGD). Сверточный слой: forward pass. Pooling, stride, варианты padding.
4. Vanishing/exploding градиенты в простой рекуррентной сети, как LSTM (частично) решает проблему.
5. Методы оптимизации SGD+momentum, RMSprop. Dropout и батч-нормализация.
6. GAN, дискриминатор/генератор, функции потерь дискриминатора/генератора. Алгоритм обучения, ошибки при обучении (коллапс моды, затухающие градиенты), оценка модели, метрика FID, приемы стабилизации обучения. Какие бывают GAN'ы.
7. Извлечение признаков CNN: ResNet (схема работы), ResNeXt (grouped convolutions).
8. Извлечение признаков CNN: DenseNet, MobileNetV1 (depthwise separable convolutions).
9. Постановки задач object detection, semantic segmentation, instance segmentation. Оценка качества. Модель Fast R-CNN, RoI pooling.
10. Модели Faster R-CNN и Mask R-CNN, RoI Align.
11. Модели YOLOv1 и SSD. Post-processing (non-maximum suppression).
12. Какие недостатки shallow graph embedding преодолевают графовые нейронные сети. Модели GCN и GAT.
13. Векторные представления слов Word2Vec Skip-gram и FastText. Модель DeepWalk для shallow graph embedding.
14. Языковая модель ELMo. Архитектура Transformer: энкодер и декодер.
15. Языковые модели BERT, GPT, GPT-2.
16. Дистилляция знаний. Подход из статьи Хинтона (softmax with temperature). Промежуточные учителя (teacher assistants). Алгоритм FitNet.

17. Абстрактное синтаксическое дерево программы. Модель Code2seq для генерации описания программного кода на естественном языке (три подзадачи из статьи).

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине (модулю)				
Оценка РО и соответствующие виды оценочных средств	2	3	4	5
Знания (виды оценочных средств: устные и письменные опросы и контрольные работы, тесты, и т.п.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (виды оценочных средств: практические контрольные задания, написание и защита рефератов на заданную тему и т.п.)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) (виды оценочных средств: выполнение и защита курсовой работы, отчет по практике, отчет по НИР и т.п.)	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

**Филиал московского государственного университета
имени М.В. Ломоносова в г. Севастополе**

Филиал московского государственного университета имени М.В. Ломоносова в
г. Севастополе

Специальность 38.03.01 Экономика

Учебная дисциплина **Многомерный статистический анализ и машинное обучение**

Семестр 7

Экзаменационный билет № 1

1. Задачи машинного обучения.

2. Оценка качества работы алгоритма.

Утверждено на заседании кафедры

Протокол № ____ от « ____ » _____ 20 ____ г.

Зав. кафедрой _____

Преподаватель _____

8. Ресурсное обеспечение:

а) Основная литература:

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511020>
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511121>
3. Орел, Е. Н. Непрерывные математические модели : учебное пособие для вузов / Е. Н. Орел, О. Е. Орел. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 120 с. — (Высшее образование). — ISBN 978-5-534-08079-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/516112>
4. Сидняев, Н. И. Теория планирования эксперимента и анализ статистических данных : учебник и практикум для вузов / Н. И. Сидняев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 495 с. — (Высшее образование). — ISBN 978-5-534-05070-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/510480>
5. Чернышев, С. А. Основы программирования на Python : учебное пособие для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/532446>

б) лицензионное программное обеспечение: нет

в) профессиональные базы данных и информационных справочных систем

1. <https://mlbootcamp.ru/> - конкурсы по машинному обучению mail.ru
2. <https://boosters.pro/> - конкурсы по машинному обучению

г) перечень ресурсов информационно-телекоммуникационной сети «Интернет»

3. <https://cran.r-project.org/> - язык R
4. <https://rstudio.com/> - оболочка Rstudio

д) Описание материально-технического обеспечения.

Для проведения семинарских занятий имеются компьютерные классы с установленным свободно распространяемым языком программирования R и оболочкой Rstudio и доступом в Интернет.

**9.Соответствие результатов обучения по данному элементу ОПОП
результатам освоения ОПОП**

-указано в общей характеристике ОПОП.

10. Язык преподавания - русский

**11.Преподаватель - кандидат физико-математических наук, доцент,
Артамонов Дмитрий Вячеславович**

**12.Автор программы - кандидат физико-математических наук, доцент,
Артамонов Дмитрий Вячеславович**