

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В.Ломоносова
Экономический факультет

УТВЕРЖДАЮ

Декан экономического факультета МГУ
профессор.

_____/ А.А.Аузан /

«__» _____ 20 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Машинное обучение и анализ данных-2

Уровень высшего образования:

магистратура

Направление подготовки (специальность):

38.04.01 ЭКОНОМИКА

Форма обучения:

Очная

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки «_____» магистратуры

ОС МГУ утвержден решением Ученого совета МГУ имени М.В.Ломоносова от _____ 20 _____ года (протокол № ____).

Год (годы) приема на обучение: 2024 и последующие

1 Место и статус дисциплины в структуре основной профессиональной образовательной программы подготовки магистра

Статус дисциплины: *обязательная (по выбору программы)*

Триместр: 3

2 Входные требования для освоения дисциплины

Для успешного освоения данного курса требуются знания и навыки, полученные в следующих дисциплинах:

- теория вероятностей и математическая статистика;
- основы программирования;
- Машинное обучение и анализ данных-1

3 Планируемые результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	Индикаторы (показатели) достижения компетенций	Планируемые результаты обучения по дисциплине (модулю), сопряженные с компетенциями
ОПК-5. Способен использовать современные информационные технологии и программные средства при решении профессиональных задач	ОПК-5.И-1. Применяет общие или специализированные пакеты прикладных программ, предназначенные для обработки, визуализации и анализа данных, в том числе эконометрического анализа и имитационного моделирования	ОПК-5.И-1.У-1. Умеет применять общие или специализированные пакеты прикладных программ (MS Excel, Eviews, Stata, SPSS, AnyLogic, Tableau и др.) или языки программирования (R, Python и др.), предназначенные для обработки, визуализации и анализа данных, в том числе эконометрического анализа и имитационного моделирования в, в соответствии со своими профессиональными задачами
ПК-8. Способен использовать различные инструментальные методы расчета и анализа	ПК-8.И-2. Применяет современные инструментальные методы расчета и анализа	ПК-8.И-2.З-1. Знает платформы анализа данных, подходящие для реализации выбранного метода расчета и

социально-экономических показателей	социально-экономических показателей при решении практических и (или) исследовательских задач	анализа, в том числе платформы бизнес-аналитики, BI-системы, платформы визуализации данных. ПК-8.И-2.У-1. Умеет применять инструментальные методы расчета и анализа показателей, в том числе с использованием различных платформ анализа данных и с применением языков программирования R и Python. ПК-8.И-2.У-2. Умеет адаптировать существующие инструментальные методы расчета и анализа показателей под конкретную задачу, в том числе путем написания программного кода на языках R и Python.
МПК-1. Способен разрабатывать эконометрические модели и модели машинного обучения исследуемых экономических процессов и явлений, интерпретировать полученные результаты	МПК-1.И-1. Применяет продвинутый эконометрический инструментарий и методы машинного обучения для построения моделей экономических процессов и явлений. Интерпретирует результаты проведенного моделирования, делает выводы и дает рекомендации на его основе	МПК-1.И-1.У-1. Умеет разрабатывать эконометрические модели и модели, основанные на методах машинного обучения. МПК-1.И-1.У-2. Умеет интерпретировать результаты моделирования, как эконометрического, так и основанного на применении методов машинного обучения. МПК-1.И-1.У-3. Умеет делать выводы из результатов моделирования, как эконометрического, так и основанного на применении методов машинного обучения, и давать рекомендации на основе сделанных выводов.
МПК-4. Способен видеть логические связи в системе собранной, обработанной и проанализированной информации, и на основании этого разрабатывать	МПК-4.И-1. Анализирует и систематизирует собранные данные, разрабатывает рекомендации для лиц, принимающих решения на основе собранных данных	МПК-4.И-1.У-1. Умеет строить логические взаимосвязи в системе собранной информации МПК-4.И-1.У-2. Умеет разрабатывать рекомендации

рекомендации для лиц, принимающих решения на микро- и макро уровне, или бизнес-решения		для лиц, принимающих управленческие решения как сфере государственного управления на разных уровнях, так и в сфере бизнеса
--	--	--

4 Объем дисциплины по видам занятий

Общая трудоемкость дисциплины – 3 з.е. (108. ч). Аудиторная нагрузка – 52 ч. (4 ч. в неделю), из них семинары – 28 ч., групповая контактная работа – 0 ч., индивидуальная контактная работа – 24 ч. Самостоятельная работа – 56 ч. Форма промежуточной аттестации – экзамен (3 триместр).

5 **Формат обучения:** используется электронная информационная среды экономического факультета МГУ имени М.В.Ломоносова «ON.ECON».

Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий Название раздела/темы	Всего, часы	В том числе		
		Контактная работа с преподавателем		Самостоятельная работа магистранта, часы
		Семинары, часы	Групповая, часы	
Тема 1. Оптимизация и регуляризация в глубоком машинном обучении	20	4	4	12
Тема 2: Сверточные нейросети для анализа изображений	16	4	4	8
Тема 3. Задача ранжирования и рекомендательные системы	20	4	4	12
Тема 4. Глубокое обучение без учителя	16	4	4	8
Тема 5. Глубокие порождающие модели	16	4	4	8
Тема 6. Обучение с подкреплением	16	4	4	8
Промежуточная аттестация (контроль): — письменная итоговая работа	4	4	—	—
Всего	108	28	24	56

Краткое содержание тем дисциплины

Тема 1. Оптимизация и регуляризация в глубоком машинном обучении. Методы оптимизации: эвристические алгоритмы (алгоритм имитации отжига), методы первого и второго порядка для дифференцируемых функций, метод импульсов, адаптивные варианты градиентного спуска. Структура глубокой нейросети на примере многослойного перцептрона, виды функций активаций. Граф вычислений и дифференцирование на нём. Стохастический градиентный спуск. Регуляризаторы и методы инициализации весов нейронной сети. Dropout. Нормализация по мини-батчам.

Тема 2. Сверточные нейросети для анализа изображений. Типы задач: классификация, детектирование объектов, сегментация. Операция свёртки для одномерных и двумерных признаков. Суб- и супердискретизация. Операция деконволюции. Аугментация данных.

Тема 3. Задача ранжирования и рекомендательные системы. Постановка задач, метрики качества. Поточечный, попарный и списковый подходы к решению задачи ранжирования. Коллаборативная фильтрация. Матричные разложения. libFM.

Тема 4. Глубокое обучение без учителя. Автокодировщики, шумоподавляющие автокодировщики, вариационные автокодировщики, поиск аномалий.

Тема 5. Глубокие порождающие модели. Состязательные сети. Архитектуры, основанные на GAN.

Тема 6. Обучение с подкреплением. Марковские процессы принятия решения. Уравнения Беллмана. TD-обучение: модели SARSA и Q-Learning. Модели глубокого обучения: DQN & DDQN.

Фонд оценочных средств для оценивания результатов обучения по дисциплине

Шкала оценивания результатов (баллы) по дисциплине:

Результаты обучения по дисциплине	Виды оценочных средств
ОПК-5.И-1.У-1. Умеет применять общие или специализированные пакеты прикладных программ (MS Excel, Eviews, Stata, SPSS, AnyLogic, Tableau и др.) или языки программирования (R, Python и др.), предназначенные для обработки, визуализации и анализа данных, в том числе эконометрического анализа и имитационного моделирования в, в соответствии со своими профессиональными задачами	Домашние работы Выполнение проектных заданий
ПК-8.И-2.3-1. Знает платформы анализа данных, подходящие для реализации выбранного метода расчета и анализа, в том числе платформы бизнес-аналитики, BI-системы, платформы визуализации данных.	Домашние работы Выполнение проектных заданий

ПК-8.И-2.У-1. Умеет применять инструментальные методы расчета и анализа показателей, в том числе с использованием различных платформ анализа данных и с применением языков программирования R и Python.	Домашние работы Выполнение проектных заданий
ПК-8.И-2.У-2. Умеет адаптировать существующие инструментальные методы расчета и анализа показателей под конкретную задачу, в том числе путем написания программного кода на языках R и Python.	Домашние работы Выполнение проектных заданий
МПК-1.И-1.У-1. Умеет разрабатывать эконометрические модели и модели, основанные на методах машинного обучения.	Домашние работы Выполнение проектных заданий
МПК-1.И-1.У-2. Умеет интерпретировать результаты моделирования, как эконометрического, так и основанного на применении методов машинного обучения.	Домашние работы Выполнение проектных заданий
МПК-1.И-1.У-3. Умеет делать выводы из результатов моделирования, как эконометрического, так и основанного на применении методов машинного обучения, и давать рекомендации на основе сделанных выводов.	Домашние работы Выполнение проектных заданий
МПК-4.И-1.У-1. Умеет строить логические взаимосвязи в системе собранной информации.	Домашние работы Выполнение проектных заданий
МПК-4.И-1.У-2. Умеет разрабатывать рекомендации для лиц, принимающих управленческие решения как в сфере государственного управления на разных уровнях, так и в сфере бизнеса	Домашние работы Выполнение проектных заданий

Виды оценочных средств	Баллы
Домашние работы	120
Письменная работа (экзамен)	30

Оценка по дисциплине выставляется, исходя из следующих критериев:

Оценка	Минимальное количество баллов	Максимальное количество баллов
<i>Отлично</i>	127,5	150

<i>Хорошо</i>	97,5	127
<i>Удовлетворительно</i>	60	97
<i>Неудовлетворительно</i>	0	59

Примечание: в случае, если магистрант за триместр набирает менее 20% баллов от максимального количества по дисциплине, то уже на промежуточном контроле (и далее на пересдачах) действует следующее правило сдачи: «магистрант может получить только оценку «Удовлетворительно», и только если получит за промежуточный контроль, включающий весь материал дисциплины, не менее, чем 85% от баллов за промежуточный контроль».

Типовые задания, методические рекомендации по их подготовке и требования к их выполнению:

Домашние работы представляют собой практические задания, ориентированные на закрепление навыков машинного обучения и анализа данных на языке Python в рамках различных тем. Задание сдается в формате Jupyter-ноутбука .ipynb с кодом, комментариями и ответами на вопросы задания. В комментариях следует описать используемые для решения задачи методы и данные, подробно обосновать выбор алгоритма, представить результаты расчётов (при необходимости используя таблицы и рисунки). Домашние задания позволяют набрать до 120 баллов.

Экзамен проходит письменно, выполняется индивидуально в день в соответствии с расписанием. Может включать теоретические задания, а также доказательства, выводы формул и расчётные задачи.

Ниже приведены примеры домашних заданий, которые будут задаваться студентам в процессе обучения.

Пример домашнего задания 1: Реализовать на языке Python рассмотренные на лекции адаптивные варианты стохастического градиентного спуска и сравнить их качество на задаче поиска минимума конкретной функции. Оформить эту работу в среде Jupyter Notebook.

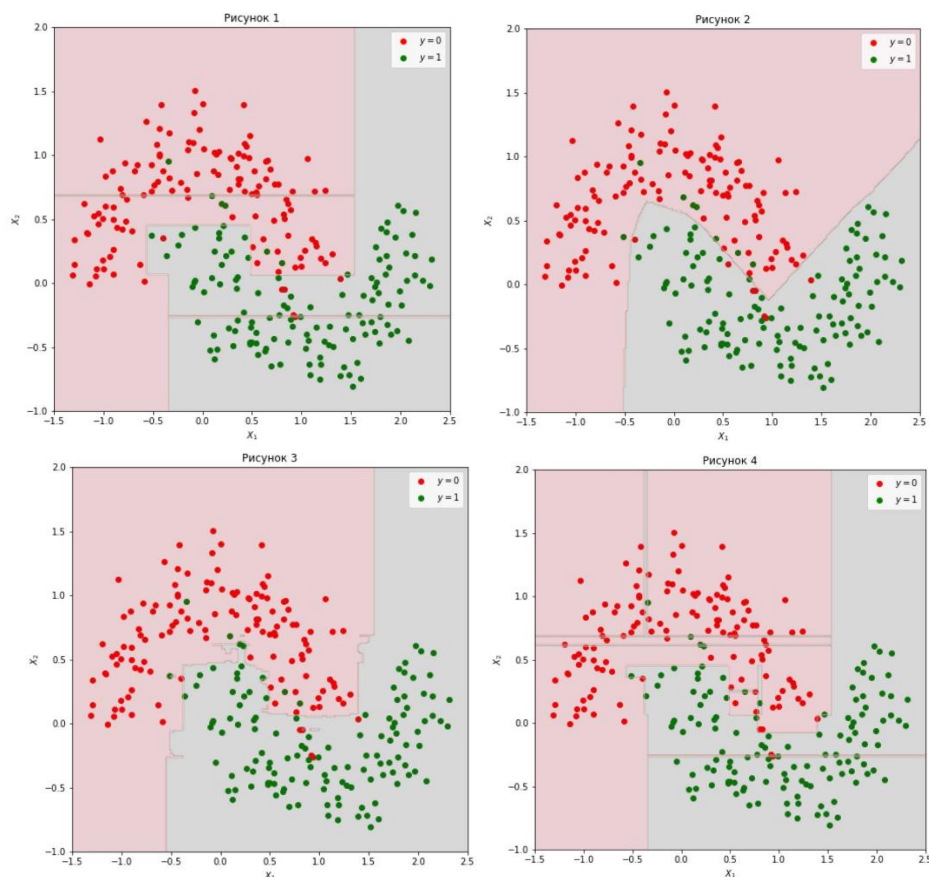
Пример домашнего задания 2: На тренировочной выборке обучить свёрточную нейронную сеть сегментации изображений земли, полученных со спутников. Обученную модель верифицировать на тестовом изображении. Результат оформить в среде Jupyter Notebook.

Пример домашнего задания 3: С помощью скрытой марковской модели и алгоритма Баума-Велша расшифровать текст, зашифрованный простым подстановочным шифром. Результат оформить в среде Jupyter Notebook.

Пример задания из письменного экзамена:

Вопрос 5

На рисунках 1-4 изображены результаты классификации. Точками обозначена обучающая выборка, цветом указаны области, которые классифицируются в класс 0 (красные) и в класс 1 (зеленые). Выберите рисунки, результат на которых может быть получен обучением решающего дерева:



8 Ресурсное обеспечение

8.1 Перечень основной и дополнительной литературы

Основная литература:

- Christopher M. Bishop. Pattern Recognition and Machine Learning
- Kevin P. Murphy. Machine Learning: A Probabilistic Perspective
- Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning
- Richard S. Sutton, Andrew G. Barto. Reinforcement Learning: An Introduction
- Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow

8.2 Перечень лицензионного программного обеспечения

- Среда разработки Anaconda для языка Python.
- Библиотеки анализа данных для языка Python.

8.3 Перечень профессиональных баз данных и информационных справочных систем

- Школа Анализа Данных Яндекса.

8.4 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

- Курс CS229 Stanford_
- Курс CS231N Stanford_
- Курс CS234 Stanford_

- Towards Data Science₂
- Платформа Kaggle₂
- Блог А.Г. Дьяконова₂
- Курс Open Data Science₂

8.5. Описание материально-технической базы

Для организации занятий по дисциплине необходимы следующие технические средства обучения: компьютерный класс с проектором и доской.

1 Язык преподавания: русский.

2 Преподаватели:

Илишаев Семен Исроильевич, руководитель направления, Блок Риски, ПАО Сбербанк

Кирпа Вадим Дмитриевич, руководитель направления по исследованию данных в Блоке Риски ПАО Сбербанк

Никонов Максим Викторович, глава аналитики и больших данных, VK.

Авторы:

Илишаев Семен Исроильевич, руководитель направления, Блок Риски, ПАО Сбербанк

Кирпа Вадим Дмитриевич, руководитель направления по исследованию данных в Блоке Риски ПАО Сбербанк

Никонов Максим Викторович, глава аналитики и больших данных, VK.

Директор магистратуры, к.э.н., доцент Е. Н. Кудряшова _____ (подпись)

Подписи преподавателей: