

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

УТВЕРЖДАЮ
декан факультета вычислительной
математики и кибернетики


/И.А. Соколов /
«27» сентября 2022г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

Практикум по методам машинного обучения

Уровень высшего образования:

бакалавриат

Направление подготовки / специальность:

01.03.02 "Прикладная математика и информатика" (3++)

Направленность (профиль) ОПОП:

Искусственный интеллект и анализ данных

Форма обучения:

очная

Рассмотрен и утвержден

на заседании Ученого совета факультета ВМК

(протокол №7, от 27 сентября 2022 года)

Москва 2022

1. ФОРМЫ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В процессе и по завершении изучения дисциплины оценивается формирование у студентов следующих компетенций:

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-5. Способен использовать инструментальные средства для решения задач машинного обучения	ПК-5.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи ПК-5.2. Разрабатывает модели машинного обучения для решения задач ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения	Знать основные современные математические модели и алгоритмы машинного обучения Уметь применять современные методы машинного обучения для решения прикладных задач анализа и обработки данных Владеть навыками реализации программных решений прикладных задач анализа и обработки данных на языке Python с использованием библиотек машинного обучения

1.1. Текущий контроль успеваемости

Текущий контроль успеваемости осуществляется путем оценки результатов выполнения заданий практических (семинарских) занятий, самостоятельной работы, предусмотренных учебным планом и посещения занятий/активность на занятиях.

В качестве оценочных средств текущего контроля успеваемости предусмотрены:

решение индивидуальных заданий

Примеры заданий

- 1) Применить алгоритмы линейной регрессии, логистической регрессии и метода опорных векторов для модельных задач.
- 2) Реализовать алгоритмы анализа данных на основе композиции классификаторов и отбора признаков
- 3) Реализовать алгоритм обратного распространения ошибки и обучение нейросети на задаче распознавания рукописных цифр MNIST.
- 4) Решить задачу предсказания опасных событий для страховой компании
- 5) Решить задачу на поиск аномалий в данных
- 6) Показатель X в классах $K1$ и $K2$ распределён нормально с параметрами: в $K1$ математическое ожидание 2, стандартное отклонение 4; в $K2$ математическое ожидание 3, стандартное отклонение 1. Выделить на числовой оси значений показателя X области отнесения байесовским классификатором к классам $K1$ и $K2$. Априорные вероятности классов $K1$ и $K2$ равны 0.6 и 0.4 соответственно.

- 7) Каждый год варан подрастает на $A\%$ от своего веса в начале года. A – случайная величина с известными матожиданием 5 и дисперсией 1 (одна и та же для всех варанов во все годы). В начале жизни каждый варан имеет вес 1. Построить байесовский классификатор для определения возраста варана (полных лет) по его весу, минимизирующий частоту ошибки. Предположить, что распознаваться будут «достаточно» взрослые вараны.
- 8) Выборка объектов из класса 1 и класса 2 определяется таблицами ниже. Указать тупиковые тесты.

	X1	X2	X3	X4						X1	X2	X3	X4	
Об.1	0	1	1	0						Об.1	0	1	0	0
Об.2	0	0	1	1						Об.2	1	0	1	0
Об.3	1	0	0	1						Об.3	0	1	0	0
Об.4	1	0	1	1						Об.4	1	1	0	0

- 9) Тестирование в банке системы распознавания для определения недобросовестных заёмщиков выявило связь между чувствительностью и ложной тревогой, показанную в таблице. Определить, приведёт ли эксплуатация системы к увеличению доходов банка. Определить возможный прирост дохода в расчёте на одну поданную заявку. Известно, что доход банка на одного заёмщика составляет 40000 денежных единиц, потери в результате отказа заёмщика от платежей составляют 120000 единиц. Доля недобросовестных заёмщиков составляет 7%.

Чувст.	Лож. Тр.
0.02	0.0001
0.12	0.003
0.23	0.05
0.38	0.12
0.47	0.16
0.58	0.19
0.67	0.23
0.78	0.34
0.89	0.52
0.97	0.72
1	0.87

- 10) В таблице даны значения переменных X и Y для четырёх экспериментов. Найти коэффициент корреляции и значения коэффициентов a и b для оптимальной по методу наименьших квадратов линейной модели $Y=a+b \cdot X$.

X	Y
0.12	52
0.23	37
0.35	17
0.46	2

- 11) Рассматривается задача классификации на два класса: положительный и отрицательный. В ходе тестирования классификатора получены следующие результаты: полнота составляет 75%, общая точность составляет 80%. Какие значения может принимать точность?
- 12) Магазин собрал сведения о покупках (транзакции в файле). Были построены ассоциативные правила. Какое правило, содержащее в условии 2 элемента, имеет наибольшую поддержку?
- 13) Государственная избирательная комиссия зафиксировала результаты выборов по партиям и по регионам (таблица в файле). Требуется кластеризовать регионы по правилу k -средних для числа кластеров K от 1 до 12. Для каждого числа кластеров K найти максимальный радиус кластера. Построить график этой величины от K . На основании графика предположить, сколько групп регионов разумно выделить по итогам выборов.

- 14) В алгоритме вычисления оценок написать формулу для числа голосов, если система опорных множеств состоит из всех непустых подмножеств, а функция близости определяется только порогами e_1, \dots, e_n .
- 15) Обоснуйте способ построения всех тупиковых тестов через приведение системы тестовых уравнений к неупрощаемой ДНФ.

1.2. Промежуточная аттестация

Промежуточная аттестация осуществляется в форме зачета

В качестве средств, используемых на промежуточной аттестации предусматривается:

Билеты

1.3. Типовые задания для проведения промежуточной аттестации

Вопросы к зачету

- 1) Вывод алгоритма xgBoost. Чем он превосходит градиентный бустинг?
- 2) Задача отбора признаков. Отбор с помощью важности признаков.
- 3) Байесовское решающее правило минимальной цены и его упрощения.
- 4) Генеративные и дискриминативные модели.
- 5) Коллаборативная фильтрация.
- 6) Основные методы кластеризации.
- 7) Методы активного обучения.
- 8) Вывод EM-алгоритма.
- 9) Обнаружение аномалий методами одноклассового SVM.
- 10) Статистическая теория обучения.

2. КРИТЕРИИ ОЦЕНКИ ПО ДИСЦИПЛИНЕ

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
виды оценочных средств				
Знания (виды оценочных средств: приведены в п. 1.2.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (виды оценочных средств: приведены в п. 1.2.)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности не принципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) (виды оценочных средств: приведены в п. 1.2..)	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