

МИНПРОСВЕЩЕНИЯ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования  
"Тульский государственный педагогический университет им. Л.Н. Толстого"  
(ФГБОУ ВО "ТГПУ им. Л.Н. Толстого")

## Введение в машинное обучение и анализ данных

### рабочая программа дисциплины (модуля)

Закреплена за кафедрой	кафедра алгебры, математического анализа и геометрии
ОПОП	Направление 02.03.01 Математика и компьютерные науки направленность (профиль) Математические основы компьютерных наук
Квалификация	Бакалавр
Год начала подготовки	2022
Форма обучения	очная
Общая трудоемкость	3 з.е.

Виды контроля по семестрам:  
зачет 5

Семестр(Курс.Номер семестра на курсе)	5(3.1)		Итого	
	УП	РПД	УП	РПД
Лекции	18	18	18	18
Лабораторные	24	24	24	24
Итого ауд.	42	42	42	42
КСР	2	2	2	2
Контактная работа	44	44	44	44
Сам. работа	64	64	64	64
Часы на контроль	0	0	0	0
Практическая подготовка	0	0	0	0
Семинары	0	0	0	0
Консультации	0	0	0	0
Итого трудоемкость в часах	108	108	108	108

Программу составил(и):

*нет, ст. преподаватель, Родионов А. В.*

Рабочая программа дисциплины

**Введение в машинное обучение и анализ данных**

разработана в соответствии с ФГОС:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 02.03.01 Математика и компьютерные науки (приказ Минобрнауки России от 23.08.2017 г. № 807)

составлена на основании учебного плана:

Направление 02.03.01 Математика и компьютерные науки  
направленность (профиль) Математические основы компьютерных наук  
утвержденного Учёным советом вуза от 28.02.2022 протокол № 3.

РПД утверждена Учёным советом университета  
протокол от 28.2.2022 г. № 3

## 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины заключается в ознакомлении с базовыми понятиями машинного обучения, с основными алгоритмами машинного обучения, особенностями их применения.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:	Б1.В.ДЭ.02
<b>2.1</b>	<b>Требования к предварительной подготовке обучающегося:</b>
1.	Методы и технологии программирования
2.	Программирование
3.	Архитектура вычислительных систем
4.	вычислительная практика
<b>2.2</b>	<b>Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:</b>
1.	Компьютерная геометрия и геометрическое моделирование
2.	Криптография и кодирование
3.	практика по получению профессиональных умений и опыта профессиональной деятельности (в том числе педагогическая практика)
4.	Моделирование бизнес-процессов
5.	научно-исследовательская работа
6.	Научные основы курса элементарной алгебры
7.	Параллельное программирование
8.	Искусственный интеллект и логическое программирование
9.	преддипломная практика

## 3. СООТНЕСЕНИЕ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ) С ИНДИКАТОРАМИ ДОСТИЖЕНИЯ КОМПЕТЕНЦИЙ

### 3.1 Компетенции обучающегося и индикаторы их достижения:

ОПК-4: Способен находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем

ОПК-4.1	Знает основные понятия, гипотезы, теоремы, методы, математические и алгоритмические модели, составляющие содержание фундаментальной и прикладной математики и связанные с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности
	- возможности алгоритмов машинного обучения; - классы задач, решаемых с помощью алгоритмов машинного обучения.
ОПК-4.2	Умеет осуществлять поиск, анализ и программную реализацию математических алгоритмов
	- применять на практике алгоритмы машинного обучения; - обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи.
ОПК-4.3	Владеет навыками программной реализации математических алгоритмов с применением современных вычислительных систем
	- базовым инструментарием машинного обучения; - программно реализовывать алгоритмы машинного обучения; - применять алгоритмы машинного обучения на практике; - анализировать результаты обучения алгоритма, предлагать пути повышения точности алгоритма.
ПК-1: Способен демонстрировать базовые знания математических и естественных наук, программирования и информационных технологий	
ПК-1.1	Обладает базовыми знаниями в области математических и естественных наук, программирования и информационных технологий
	- возможности алгоритмов машинного обучения; - классы задач, решаемых с помощью алгоритмов машинного обучения.
ПК-1.2	Умеет осуществлять поиск, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности с использованием математического моделирования процессов и объектов
	- применять на практике алгоритмы машинного обучения; - обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи.
ПК-1.3	Имеет практический опыт научно-исследовательской деятельности в областях, использующих математические методы и компьютерные технологии

- базовым инструментарием машинного обучения;
- программно реализовывать алгоритмы машинного обучения;
- применять алгоритмы машинного обучения на практике;
- анализировать результаты обучения алгоритма, предлагать пути повышения точности алгоритма.

**3.2 Результаты обучения по дисциплине:****В результате освоения дисциплины обучающийся должен:**

	<b>Знать:</b>
3.1	- возможности алгоритмов машинного обучения;
3.2	- классы задач, решаемых с помощью алгоритмов машинного обучения.
	<b>Уметь:</b>
У.1	- применять на практике алгоритмы машинного обучения;
У.2	- обосновать применение того или иного алгоритма машинного обучения для решения
У.3	конкретной задачи.
	<b>Владеть:</b>
В.1	- базовым инструментарием машинного обучения;
В.2	- программно реализовывать алгоритмы машинного обучения;
В.3	- применять алгоритмы машинного обучения на практике;
В.4	- анализировать результаты обучения алгоритма, предлагать пути повышения точности
В.5	алгоритма.

**4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)**

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Литература	Содержание
	<b>Типы задач. Метрические классификаторы. Алгоритмы кластеризации</b>				
1.1	Типы задач машинного обучения /Лек/	5	2	Л1.1 Л1.2	
1.2	Метрические классификаторы /Лек/	5	2	Л1.1 Л1.2	
1.3	Метрические классификаторы /Лаб/	5	4	Л1.1 Л1.2	
1.4	Алгоритмы кластеризации /Лек/	5	2	Л1.1 Л1.2	
1.5	Алгоритмы кластеризации /Лаб/	5	4	Л1.1 Л1.2	
1.6	Метрические классификаторы. Алгоритмы кластеризации /Ср/	5	20	Л1.1 Л1.2	
	<b>Деревья решений, линейные классификаторы. Нейронные сети</b>				
2.1	Деревья решений /Лек/	5	2	Л1.1 Л1.2	
2.2	Деревья решений /Лаб/	5	2	Л1.1 Л1.2	
2.3	Линейные классификаторы /Лек/	5	2	Л1.1 Л1.2	
2.4	Линейные классификаторы /Лаб/	5	2	Л1.1 Л1.2	
2.5	Нейронные сети /Лек/	5	2	Л1.1 Л1.2	
2.6	Нейронные сети /Лаб/	5	4	Л1.1 Л1.2	
2.7	Деревья решений, линейные классификаторы. Нейронные сети /Ср/	5	22	Л1.1 Л1.2	

	<b>Регрессионный анализ, Ансамблевые методы. Стохастический поиск</b>				
3.1	Регрессионный анализ /Лек/	5	2	Л1.1 Л1.2	
3.2	Регрессионный анализ /Лаб/	5	2	Л1.1 Л1.2	
3.3	Ансамблевые методы /Лек/	5	2	Л1.1 Л1.2	
3.4	Ансамблевые методы /Лаб/	5	2	Л1.1 Л1.2	
3.5	Стохастический поиск /Лек/	5	2	Л1.1 Л1.2	
3.6	Стохастический поиск /Лаб/	5	4	Л1.1 Л1.2	
3.7	КСР /КСР/	5	2	Л1.1 Л1.2	
3.8	Регрессионный анализ, Ансамблевые методы. Стохастический поиск /Ср/	5	22	Л1.1 Л1.2	

## 5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

### 5.1. Типовые задания для проведения текущего контроля

Пример задания №1:

Задача 1. [3 балла] Реализуйте алгоритм kNN классификации по k ближайшим соседям, используя простое евклидовое расстояние.

Задача 2. [3 балла] Реализуйте алгоритм k-means для кластеризации на 2-4 кластера.

Задача 3. [4 балла] Реализуйте алгоритм DBSCAN, найдите параметры для кластеризации на 4 кластера.

Пример задания №2:

Задача 1. [2,5 балла] Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10).

Задача 2. [2,5 балла] Примените метод SVM (например, из библиотеки sklearn) для датасета blobs2. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF).

Задача 3. [2,5 балла] Реализуйте алгоритм логистической регрессии со стохастическим градиентным спуском, обучите его на датасете spambase\_old (train) и проверьте на датасете spambase\_new (val). Получите ROC кривые для вариантов без нормировки и с нормировкой признаков.

Задача 4. [2,5 балла] Модифицируйте модель из задачи 3, заменив последний нейрон на 10 нейронов, и реализовав мультиклассовую классификацию с softmax в качестве решающей функции и кросс-энтропией в качестве функции потерь и обучите на подготовленном датасете mnist.

Пример задания №3:

Задача 1. [2,5 балла] Реализуйте алгоритм линейной регрессии, и полиномиальной регрессии (для датасета poisysine – степеней от 2 до 5, для датасета hydrodynamics – степени 2) без регуляризации.

Задача 2. [2,5 балла] Реализуйте алгоритм гребневой регрессии и найдите оптимальный параметр регуляризации для случаев из задачи 1.

Задача 3. [2,5 балла] Найдите максимум функции с помощью алгоритма кросс-энтропийного поиска, изображая распределение на каждом шаге.

Задача 4. [2,5 балла] Найдите лучший путь в задаче коммивояжера с помощью алгоритма отжига

### 5.2. Типовые задания для проведения промежуточной аттестации

1. Препроцессинг. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding.
2. Кластеризация. kMeans, MeanShift, DBSCAN, Affinity Propagation.
3. Смещение и дисперсия (bias and variance). Понятие средней гипотезы.
4. Ансамблевые методы. Soft and Hard Voting. Bagging. Случайные леса. AdaBoost.
5. Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение.
6. Бустинг деревьев решений.
7. Ошибка внутри и вне выборки. Ошибка обобщения. Неравенство Хёфдинга. Валидация и кросс-валидация.
8. Линейная регрессия. Полиномиальная регрессия. Гребневая регрессия.

9. Размерность Вапника-Червоненкиса. Размерность Вапника-Червоненкиса для перцептрона.
10. Логистическая регрессия. Градиентный спуск.
11. Пороговые условия. Эффективность по Парето. Precision-Recall и ROC кривые. AUC.
12. Ансамблевые методы регрессии. RANSAC. Theil-Sen. Huber.
13. Перцептрон. Перцептрон с карманом.
14. Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра.
15. Гипотезы и дихотомии. Функция роста. Точка поломки. Доказательство полиномиальности функции роста в присутствии точки поломки.
16. Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев. Небрежные решающие деревья.
17. Байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). EM алгоритм.
18. Нейронные сети. Перцептрон Розенблатта. Функции активации. Обратное распространение градиента. Softmax.
19. Стохастическая оптимизация. Hill Climb. Отжиг. Генетический алгоритм.
20. Метрические классификаторы. kNN. WkNN. Отбор эталонов. DROP5. Kdtree

### 5.3. Перечень видов оценочных средств

Зачет  
Лабораторные работы  
Индивидуальные задания

### 5.4. Процедура применения оценочных материалов

Итоговая рейтинговая оценка по дисциплине складывается из следующих составляющих:  
Зачет 30 баллов  
Лабораторные работы 50 баллов  
Индивидуальные задания 20 баллов

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

### 6.1. Рекомендуемая литература

#### 6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год (кол-во экземпляров для печатных изданий)	Ссылка на электронное издание
Л1.1	Барский А. Б.	Логические нейронные сети: учебное пособие	, 2007	<a href="http://biblioclub.ru/index.php?page=book&amp;id=232983">http://biblioclub.ru/index.php?page=book&amp;id=232983</a>
Л1.2	Яхьяева, Г. Э.	Нечеткие множества и нейронные сети: учебное пособие	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020	<a href="http://www.iprbookshop.ru/97552.html">http://www.iprbookshop.ru/97552.html</a>

### 6.3. Информационные технологии

#### 6.3.1 Перечень лицензионного и свободно распространяемого программного обеспечения

1.	Операционная система Microsoft Windows 10 Professional Russian. Контракт № ПР/ФЕН/15/18 от 23.10.2015 г., договор № Пр/16/6 от 05 апреля 2016 г.
2.	Программное обеспечение Microsoft Office Enterprise 2007 Russian. Лицензия №46138962 от 16.11.2009
3.	Браузеры Google Chrome, Mozilla, Opera. Свободно распространяемое ПО
4.	Система Интернет-телефонии Skype. Свободно распространяемое ПО

#### 6.3.2 Перечень информационных справочных систем и профессиональных баз данных

1.	Национальная энциклопедическая служба ( <a href="https://vocabulary.ru">https://vocabulary.ru</a> )
2.	Библиотека федерального портала «Российское образование» ( <a href="http://www.edu.ru">http://www.edu.ru</a> )
3.	Web of Science Core Collection – политематическая реферативно-библиографическая и наукометрическая (библиометрическая) база данных ( <a href="http://webofscience.com">http://webofscience.com</a> )
4.	Портал «Информационно-коммуникационные технологии в образовании» ( <a href="http://www.ict.edu.ru">http://www.ict.edu.ru</a> )

## 7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Ауд.	Назначение	Оборудование и технические средства обучения	Вид
4-302	Учебная аудитория	доска учебная, проектор, стол преподавателя, столы учебные, стул преподавателя, экран	

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

Самостоятельная работа может рассматриваться как организационная форма обучения – система педагогических условий, обеспечивающих управление учебной деятельностью по освоению знаний и умений в области учебной деятельности без посторонней помощи. Студенту нужно четко понимать, что самостоятельная работа – не просто обязательное, а необходимое условие для получения знаний по дисциплине и развитию компетенций, необходимых в будущей профессиональной деятельности.

Самостоятельная работа проводится с целью:

- систематизации и закрепления полученных на лекциях теоретических знаний;
- углубления и расширения теоретических знаний;
- формирования умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развития познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- формирования практических (общеучебных и профессиональных) умений и навыков;
- развития исследовательских умений;
- получения навыков эффективной самостоятельной профессиональной (практической и научно-теоретической) деятельности.

В учебном процессе выделяют два вида самостоятельной работы:

- аудиторная;
- внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа - планируемая учебная работа студентов, выполняемая во внеаудиторное время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Самостоятельная работа, не предусмотренная программой учебной дисциплины, раскрывающей и конкретизирующей ее содержание, осуществляется студентом инициативно, с целью реализации собственных учебных и научных интересов.

Для более эффективного выполнения самостоятельной работы по дисциплине преподаватель рекомендует источники для работы, характеризует наиболее рациональную методику самостоятельной работы, демонстрирует ранее выполненные студентами работы и т. п.

Виды заданий для внеаудиторной самостоятельной работы, их содержание и характер могут иметь вариативный и дифференцированный характер, учитывать индивидуальные особенности студента.

Самостоятельная работа может осуществляться индивидуально или группами студентов online и на занятиях в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности.

Контроль результатов внеаудиторной самостоятельной работы осуществляется в пределах времени, отведенного на обязательные учебные занятия по дисциплине на практических занятиях.

Для представления результатов самостоятельной работы рекомендуется:

Составить план выступления, в котором отразить тему, самостоятельный характер проделанной работы, главные выводы и/или предложения, их краткое обоснование и практическое значение – с тем, чтобы в течение 3 – 5 минут представить достоинства, выполненного самостоятельно задания.

Подготовить иллюстративный материал в виде презентации для использования во время представления результатов самостоятельной работы в аудитории. Конкретный вариант наглядного представления результатов определяется форматом аудиторного занятия и задания преподавателя.