

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Физический факультет



УТВЕРЖДАЮ:  
декан физического факультета

С.Н. Филимонов

2021 г.

Рабочая программа дисциплины

**Основы машинного обучения**

по направлению подготовки

**03.03.02 Физика**

Направленность (профиль) подготовки:  
**«Фундаментальная физика»**

Форма обучения  
**Очная**

Квалификация  
**Бакалавр**

Год приема  
**2021**

Код дисциплины в учебном плане: Б1.В.ДВ.01.08.15

СОГЛАСОВАНО:

Руководитель ОП

 О.Н. Чайковская

Председатель УМК

 О.М. Сюсина

Томск – 2021

## **1. Цель и планируемые результаты освоения дисциплины (модуля)**

Целью освоения дисциплины является формирование следующих компетенций:

- УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач;
- ОПК-3 – Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности.
- ПК-1 – Способен проводить научные исследования в выбранной области с использованием современных экспериментальных и теоретических методов, а также информационных технологий.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

- ИУК 1.1. Осуществляет поиск информации, необходимой для решения задачи;
- ИУК 1.2. Проводит критический анализ различных источников информации (эмпирической, теоретической);
- ИУК 1.3. Выявляет соотношение части и целого, их взаимосвязь, а также взаимоподчиненность элементов системы в ходе решения поставленной задачи;
- ИОПК 3.1. Знает основы программирования и требования информационной безопасности;
- ИОПК 3.2. Применяет общее и специализированное программное обеспечение для теоретических расчетов и обработки экспериментальных данных;
- ИПК 1.2. Владеет практическими навыками использования современных методов исследования в выбранной области.

## **2. Задачи освоения дисциплины**

- Освоить теоретические основы машинного обучения для решения прикладных задач, в том числе и на реальных данных.
- Научиться применять понятийную базу и передовыми инструментами машинного обучения для решения научных и практических задач.

## **3. Место дисциплины (модуля) в структуре образовательной программы**

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

## **4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине**

Семестр 8, зачет с оценкой.

## **5. Входные требования для освоения дисциплины**

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: Основы программирования на языке Python.

## **6. Язык реализации**

Русский

## **7. Объем дисциплины (модуля)**

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

- лекции: 24 ч.;
- практические занятия: 24 ч.;

Объем самостоятельной работы студента определен учебным планом.

## **8. Содержание дисциплины (модуля), структурированное по темам**

### **Тема 1. Основы Python**

Передовые инструменты машинного обучения (МО) исторически разрабатывались и разрабатываются в основном на языке Python. Данная тема является фундаментом, на котором будет строиться дальнейшая работа. Главные вопросы: основные типы и структуры данных, синтаксис, особенности языка

### **Тема 2. Библиотеки Python: NumPy**

Библиотека Numeric Python или NumPy представляет собой набор методов и типов данных, которые используются в наукоёмкой разработке. Так же эта библиотека лежит в основе других наукоёмких библиотек. Главные вопросы: основные типы данных и приёмы работы с ними; векторные операции; стандартные процедуры.

### **Тема 3. Библиотеки Python: matplotlib**

Библиотека matplotlib позволяет строить различные виды графиков и чертежей. Визуализация данных является важным этапом их анализа. Главные вопросы: построение сложных графиков и их оформление.

### **Тема 4. Библиотеки Python: pandas**

Библиотека pandas позволяет работать с табличными данными. Главные вопросы: основные типы данных и приёмы работы с ними; статистические процедуры; связь с реляционной алгеброй.

### **Тема 5. Теоретические основы машинного обучения:**

Постановка задачи МО в общем виде. Типы задач. Главные вопросы: задача; метрики качества; валидация; переобучение; обнаружение и обработка выбросов; метрические методы и гипотеза компактности; функции потерь; обучение с учителем и без.

### **Тема 6. Метрические алгоритмы машинного обучения:**

Разбор основных алгоритмов МО, с точки зрения механики их работы и границ применимости. Алгоритмы: метод k-ближайших соседей; метод опорных векторов; метод опорных векторов по комбинации базисных функций; нелинейная регрессия

### **Тема 7. Линейная регрессия:**

Разбор линейной регрессии с точки зрения МО. Постановка задачи линейной регрессии. Главные вопросы: формула в общем виде; методы поиска решений и границы применимости; линейная регрессия по комбинации базисных функций; извлечение информации из коэффициентов линейной регрессии.

### **Тема 8. Решающие деревья:**

Решающие деревья - один из самых известных и универсальных алгоритмов МО. В данной теме рассматриваются теоретические основания работы решающих деревьев, выбора предикатов ветвления, основные гиперпараметры дерева и границы применения. Главные вопросы: принципиальная схема решающего дерева; предикаты ветвления; глубина дерева; переобучение дерева; визуализация структуры дерева.

### **Тема 9. Ансамблирование алгоритмов:**

Ансамблирование алгоритмов - это использование вместо одной сильной модели комбинации большого количества слабых. Рассматриваются на примере решающих деревьев. Главные вопросы: бэггинг; бустинг; мудрость толпы; сравнение ансамблей с одной моделью.

### **Тема 10. Глубинное обучение:**

Глубинное обучение - это использование глубоких искусственных нейронных сетей (ИНС) в качестве алгоритма МО. Рассматривается программная реализация в виде библиотеки для Python под названием `pytorch`. Главные вопросы: искусственные нейроны; библиотека `pytorch`; типы слоёв; функции активации; инициализация весов; оптимизаторы; нормализация данных; проектирование архитектуры; выбор функции потерь под задачу.

### **Тема 11. Основные архитектуры искусственных нейронных сетей:**

В данной теме рассматриваются основные архитектуры, т.е. некоторые структуры слоёв, ИНС и задачи, для которых они применяются. Архитектуры: полносвязные; свёрточные; автокодировщики; рекуррентные.

### **Тема 12. Сохранение, передача и работа с предобученными моделями:**

ИНС не обязательно обучать с состояния полного хаоса. Возможно использовать предобученные модели для схожих задач. Главные вопросы: сохранение моделей; формат `.pickle`; загрузка моделей; дообучение моделей (`transfer learning`).

## **9. Текущий контроль по дисциплине**

Текущий контроль по дисциплине проводится путем решения задач.

## **10. Порядок проведения и критерии оценивания промежуточной аттестации**

Зачет с оценкой проводится в устной форме по билетам. Билет содержит теоретический вопрос. Продолжительность зачета 1,5 часа.

Примерный перечень теоретических вопросов

1. Основы Python: основные типы и структуры данных, синтаксис, особенности языка.
2. Библиотека NumPy: основные типы данных и приёмы работы с ними; векторные операции; стандартные процедуры.
3. Библиотека `matplotlib`: построение сложных графиков и их оформление.
4. Библиотека `pandas`: основные типы данных и приёмы работы с ними; статистические процедуры; связь с реляционной алгеброй.
5. Теоретические основы машинного обучения: задача; метрики качества; валидация.
6. Переобучение; обнаружение и обработка выбросов; метрические методы и гипотеза компактности.
7. Функции потерь.
8. Обучение с учителем и без.
9. Метод  $k$ -ближайших соседей.
10. Метод опорных векторов.
11. Метод опорных векторов по комбинации базисных функций.
12. Нелинейная регрессия.
13. Линейная регрессия.
14. Решающие деревья.
15. Бэггинг
16. Бустинг; мудрость толпы; сравнение ансамблей с одной моделью.
17. Глубинное обучение.
18. Основные архитектуры искусственных нейронных сетей:
19. Сохранение, передача и работа с предобученными моделями.

На зачете проверяются результаты освоения дисциплины по индикаторам ИУК 1.1; ИУК 1.2; ИУК 1.3; ИОПК 3.1; ИОПК 3.2; ИПК 1.2.

Результаты зачета с оценкой определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Текущий контроль влияет на допуск к зачету.

### **11. Учебно-методическое обеспечение**

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=24612>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) План семинарских / практических занятий по дисциплине.

г) Методические указания по проведению лабораторных работ.

д) Методические указания по организации самостоятельной работы студентов.

### **12. Перечень учебной литературы и ресурсов сети Интернет**

а) основная литература:

– Ярушкина Н. Г. Интеллектуальный анализ временных рядов: Учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. - М.: ИД ФОРУМ: ИНФРА-М, 2012. - 160 с.:

– Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6.

б) дополнительная литература:

– Технология Data Mining: Интеллектуальный анализ данных, Степанов, Роман Григорьевич, 2009г

– Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.: ил. + CD-ROM (Учебная литература для вузов)

в) ресурсы сети Интернет:

– Введение в машинное обучение. Coursera. К.В. Воронцов  
<https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie>

– NumPy Documentation <https://numpy.org/doc/>

– Pytorch Documentation <https://pytorch.org/docs/stable/index.html>

### **13. Перечень информационных технологий**

а) лицензионное и свободно распространяемое программное обеспечение:

– Python 3.7+ и модули: jupyter, numpy, matplotlib, pandas, scikit-learn, pytorch

### **14. Материально-техническое обеспечение**

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

## **15. Информация о разработчиках**

Красавин Дмитрий Сергеевич, Томский государственный университет, м.н.с.

Галушина Татьяна Юрьевна, к.ф.-м.н., Томский государственный университет,  
доцент