

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное автономное образовательное учреждение
высшего образования
«Новосибирский национальный исследовательский государственный университет»
(Новосибирский государственный университет, НГУ)

**Физический факультет
Кафедра автоматизации физико-технических исследований**



УТВЕРЖДАЮ
Декан ФФ, д.ф.-м.н
В.Е.Блинов
2022 г.

Рабочая программа дисциплины

ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ

Направление подготовки **03.03.02 Физика**
Направленность (профиль): **Физическая информатика**

Форма обучения
Очная

| Семестр | Общий объем | Виды учебных занятий (в часах) | | | | Промежуточная аттестация (в часах) | | | | |
|---|-------------|--|----------------------|-------------------------------|--|---|--|-------|--------------------------|---------|
| | | Контактная работа обучающихся с преподавателем | | | Самостоятельная работа, не включая период сессии | Самостоятельная подготовка к промежуточной аттестации | Контактная работа обучающихся с преподавателем | | | |
| | | Лекции | Лабораторные занятия | Консультации в период занятий | | | Консультации | Зачет | Дифференцированный зачет | Экзамен |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 7 | 144 | 32 | 32 | | 56 | 18 | 4 | | | 2 |
| Всего 144 часа /4 зачётные единицы, из них: - контактная работа 70 часов | | | | | | | | | | |
| Компетенции ОПК-3 | | | | | | | | | | |

Ответственный за образовательную программу
д.ф.-м.н., проф.

С. В. Цыбуля

Новосибирск, 2022

Содержание

| | |
|---|---|
| 1. Перечень планируемых результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы. | 3 |
| 2. Место дисциплины в структуре образовательной программы. | 3 |
| 3. Трудоемкость дисциплины в зачётных единицах с указанием количества академических часов, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу. | 4 |
| 4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведённого на них количества академических часов и видов учебных занятий. | 4 |
| 5. Перечень учебной литературы. | 6 |
| 6. Перечень учебно-методических материалов для самостоятельной работы. | 6 |
| 7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины. | 6 |
| 8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине. | 6 |
| 9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине. | 6 |
| 10. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации по дисциплине. | 7 |

1. Перечень планируемых результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы.

Цель курса – знакомство с современными аспектами классического машинного обучения и обзор применения методов машинного обучения для решения задач экспериментальной физики. Дисциплина нацелена на формирование общепрофессиональной компетенции ОПК-3.

| Результаты освоения образовательной программы (компетенции) | Индикаторы | Результаты обучения по дисциплине |
|--|---|--|
| <p>ОПК-3. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности</p> | <p>ОПК - 3.1. Применяет различные источники информации для решения задач профессиональной сферы деятельности.</p> <p>ОПК – 3.2. Применяет основные приемы, возможности и правила работы со стандартными и специализированными программными продуктами при решении профессиональных задач.</p> <p>ОПК – 3.3. Применяет методологию поиска научной и технической информации в сети Интернет и специализированных базах данных.</p> | <p>Знать основные алгоритмы машинного обучения, основные задачи и основные метрики качества в этих задачах, современные методы расчета объекта научного исследования, методы работы в различных операционных системах, с базами данных, с экспертными системами; современные open-source инструменты по машинному обучению.</p> <p>Уметь решать типичные задачи на основе воспроизведения стандартных алгоритмов решения, применять на практике методы для решения задач классификации и задач регрессии, сводить поставленную задачу к типичным задачам машинного обучения.</p> |

2. Место дисциплины в структуре образовательной программы.

Дисциплина реализуется в осеннем семестре 4-го курса для бакалавров, обучающихся по направлению подготовки 03.03.02 Физика. Дисциплина, в основном, связана с курсами математического цикла и опирается на освоенные при их изучении знания и умения. В результате изучения курса студенты должны овладеть принципами цифровой обработки данных, представленных в основном в виде двумерных массивов, отражающих свойства физических полей различной природы. Изучение материала курса предполагает знание студентами основ математического анализа, линейной алгебры и теории вероятностей. Для ознакомления с экспериментальной частью курса и выполнения практических заданий настоятельно рекомендуется знакомство с пакетом MATLAB версии R2016 и выше. Для облегчения освоения материала курса полезно иметь опыт работы в среде MATLAB. Успешное решение практических задач требует владения программированием на одном из языков высокого уровня.

3. Трудоёмкость дисциплины в зачётных единицах с указанием количества академических часов, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу.

| Семестр | Общий объем | Виды учебных занятий (в часах) | | | | Промежуточная аттестация (в часах) | | | | |
|---|-------------|--|----------------------|-------------------------------|--|---|--|-------|--------------------------|---------|
| | | Контактная работа обучающихся с преподавателем | | | Самостоятельная работа, не включая период сессии | Самостоятельная подготовка к промежуточной аттестации | Контактная работа обучающихся с преподавателем | | | |
| | | Лекции | Лабораторные занятия | Консультации в период занятий | | | Консультации | Зачет | Дифференцированный зачет | Экзамен |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 7 | 144 | 32 | 32 | | 56 | 18 | 4 | | | 2 |
| Всего 144 часа /4 зачётные единицы, из них: - контактная работа 70 часов | | | | | | | | | | |
| Компетенции ОПК-3 | | | | | | | | | | |

Реализация дисциплины предусматривает практическую подготовку при проведении следующих видов занятий, предусматривающих участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью: лекции, лабораторные занятия, самостоятельная работа студента, консультации, экзамен.

Программой дисциплины предусмотрены следующие виды контроля:

Текущий контроль успеваемости: задания для самостоятельного решения.

Промежуточная аттестация: экзамен

Общая трудоемкость рабочей программы дисциплины составляет 144 академических часа/4 зачётные единицы:

- занятия лекционного типа – 32 часа;
- лабораторные занятия – 32 часа;
- самостоятельная работа обучающегося в течение семестра, не включая период сессии – 56 часов;
- промежуточная аттестация (подготовка к экзамену, консультации, экзамен) – 24 часа.

Объём контактной работы обучающегося с преподавателем (занятия лекционного типа, лабораторные занятия, экзамен) составляет 70 часов.

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведённого на них количества академических часов и видов учебных занятий.

Общая трудоемкость дисциплины составляет 4 зачетные единицы, 144 академических часа.

| № п/п | Раздел дисциплины | Неделя семестра | Виды учебных занятий, включая самостоятельную работу студентов и трудоемкость (в часах) | | | | Консультации | Промежуточная |
|-------|-------------------|-----------------|---|-----------------|-----------|-----------|--------------|---------------|
| | | | Всего | Аудиторные часы | Са м. раб | Са м. наб | | |

| | | | | Лекции | Лаб. работы | Консультации в период занятий | | | | |
|----|---|-------|------------|-----------|-------------|-------------------------------|-----------|-----------|----------|----------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | Базовые понятия машинного обучения | 1 | 9 | 2 | 2 | | 5 | | | |
| 2 | Линейные алгоритмы | 2-3 | 13 | 4 | 4 | | 5 | | | |
| 3 | Непараметрические. Методы классификации и регрессии | 4-5 | 14 | 4 | 4 | | 6 | | | |
| 4 | Метод опорных векторов | 6 | 9 | 2 | 2 | | 5 | | | |
| 5 | Логические методы | 7-8 | 18 | 4 | 4 | | 10 | | | |
| 6 | Ансамбли алгоритмов | 9-10 | 13 | 4 | 4 | | 5 | | | |
| 7 | Введение в нейронные сети | 11-12 | 18 | 4 | 4 | | 10 | | | |
| 8 | Работа с признаками | 13-14 | 13 | 4 | 4 | | 5 | | | |
| 9 | Кластеризация | 15-16 | 13 | 4 | 4 | | 5 | | | |
| 10 | Самостоятельная подготовка обучающегося к экзамену | | 18 | | | | | 18 | | |
| 11 | Консультации | | 4 | | | | | | 4 | |
| 12 | Экзамен | | 2 | | | | | | | 2 |
| | Итого: | | 144 | 32 | 32 | | 56 | 18 | 4 | 2 |

Программа и основное содержание лекций (32 часа)

- Базовые понятия машинного обучения (2 часа)
- Линейные алгоритмы (4 часа)
- Непараметрические. Методы классификации и регрессии (4 часа)
- Метод опорных векторов (2 часа)
- Логические методы (4 часа)
- Ансамбли алгоритмов (4 часа)
- Введение в нейронные сети (4 часа)
- Работа с признаками (4 часа)
- Кластеризация (4 часа)

Программа и основное содержание лабораторных работ (32 часа)

- Базовые понятия машинного обучения (2 часа)
- Линейные алгоритмы (4 часа)

- Непараметрические. Методы классификации и регрессии (4 часа)
- Метод опорных векторов (2 часа)
- Логические методы (4 часа)
- Ансамбли алгоритмов (4 часа)
- Введение в нейронные сети (4 часа)
- Работа с признаками (4 часа)
- Кластеризация (4 часа)

Самостоятельная работа студентов (74 часа)

| Перечень занятий на СРС | Объем, час |
|---|------------|
| Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях. | 56 |
| Подготовка к экзамену. Повторение теоретического материала по вопросам, совпадающим с темами лекций | 18 |

5. Перечень учебной литературы.

1. Л. А. Бессонов. Теоретические основы электротехники. Электрические цепи: [учебник] / 10-е изд. М.: Гардарики, 2002. - 637 с. (82 экз.).

6. Перечень учебно-методических материалов для самостоятельной работы.

2. И. С. Гоноровский. Радиотехнические цепи и сигналы: [Учеб. для радиотехн. спец. вузов] / 2-е изд., перераб. М.: Сов. радио, 1971. - 671 с. (66 экз.).

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.

7.1 Ресурсы сети Интернет

Для освоения дисциплины используются следующие ресурсы:

- электронная информационно-образовательная среда НГУ (ЭИОС);
- образовательные интернет-порталы;
- информационно-телекоммуникационная сеть Интернет.

7.2 Современные профессиональные базы данных

Не используются.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.

8.1 Перечень программного обеспечения

Для обеспечения реализации дисциплины используется стандартный комплект программного обеспечения (ПО), включающий регулярно обновляемое лицензионное ПО Windows и MS Office, и среда разработки Microsoft Visual Studio.

Для обеспечения самостоятельной работы студентов может применяться среда программирования MATLAB или Python.

8.2 Информационные справочные системы

Не используются

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине.

Для реализации дисциплины используются специальные помещения:

1. Учебные аудитории для проведения занятий лекционного типа, лабораторных занятий и промежуточной аттестации.

2. Помещения для самостоятельной работы обучающихся.

Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду НГУ.

Материально-техническое обеспечение образовательного процесса по дисциплине для обучающихся из числа лиц с ограниченными возможностями здоровья осуществляется согласно «Порядку организации и осуществления образовательной деятельности по образовательным программам для инвалидов и лиц с ограниченными возможностями здоровья в Новосибирском государственном университете».

10. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации по дисциплине.

10.1. Порядок проведения текущего контроля и промежуточной аттестации успеваемости по дисциплине.

Текущий контроль успеваемости

Текущий контроль успеваемости осуществляется в ходе семестра путем проверки заданий для самостоятельного решения. Примеры заданий приведены в п. 10.3.

Промежуточная аттестация

Освоение компетенций оценивается согласно шкале оценки уровня сформированности компетенции. Положительная оценка по дисциплине выставляется в том случае, если заявленная компетенция ОПК-3 сформирована не ниже порогового уровня.

Окончательная оценка работы студента в течение семестра происходит на экзамене. Экзамен проводится в конце семестра в устной форме. Вопросы подбираются таким образом, чтобы проверить уровень сформированности компетенции ОПК-3.

Итоговая оценка по курсу выставляется после устного экзамена с учётом оценок за работу в семестре (РС). Итоговая оценка по промежуточной аттестации не может превышать более, чем на 1 балл, оценку за работу в семестре.

Оценка за РС учитывает активность студента на лабораторных занятиях, оцениваемую преподавателем, посещение, количество сданных лабораторных работ. Для получения за РС оценки «удовлетворительно» либо выше необходимо сдать все лабораторные работы (темы работ указаны в п. 4).

Промежуточная аттестация по дисциплине проводится в экзаменационную сессию по билетам. Для получения на экзамене оценки «удовлетворительно» либо выше, необходимо предварительно сдать все лабораторные работы из каждого блока и получить оценку «удовлетворительно» за тестирование. Студент, не выполнивший эти условия в течение семестра, начинает сдачу экзамена со сдачи лабораторных работ. На это ему отводится 1 час. Если обучающийся не справляется со сдачей лабораторных работ и/или тестированием за отведенное время, ему выставляется оценка «неудовлетворительно» по итогам промежуточной аттестации. Студенты, справившиеся со сдачей лабораторных работ, приступают к подготовке к ответам по билету. На подготовку им отводится 1 час. Билеты включают два вопроса.

Вывод об уровне сформированности компетенции принимается преподавателем. Каждый вопрос билета оценивается от 0 до 5 баллов. Положительная оценка ставится, когда компетенция освоена не ниже порогового уровня. Оценки «отлично», «хорошо», «удовлетворительно» означают успешное прохождение промежуточной аттестации.

Соответствие индикаторов и результатов освоения дисциплины

Таблица 10.1

| Индикатор | Результат обучения по дисциплине | Оценочные средства |
|---|---|--|
| <p>ОПК - 3.1. Применяет различные источники информации для решения задач профессиональной сферы деятельности.</p> <p>ОПК – 3.2. Применяет основные приемы, возможности и правила работы со стандартными и специализированными программными продуктами при решении профессиональных задач.</p> <p>ОПК – 3.3. Применяет методологию поиска научной и технической информации в сети Интернет и специализированных базах данных.</p> | <p>Знать основные алгоритмы машинного обучения, основные задачи и основные метрики качества в этих задачах, современные методы расчета объекта научного исследования, методы работы в различных операционных системах, с базами данных, с экспертными системами; современные open-source инструменты по машинному обучению.</p> | <p>Задания для самостоятельного решения, экзамен в устной форме.</p> |
| | <p>Уметь решать типичные задачи на основе воспроизведения стандартных алгоритмов решения, применять на практике методы для решения задач классификации и задач регрессии, сводить поставленную задачу к типичным задачам машинного обучения.</p> | <p>Задания для самостоятельного решения, экзамен в устной форме.</p> |

10.2. Описание критериев и шкал оценивания индикаторов достижения результатов обучения по дисциплине «Введение в машинное обучение».

Таблица 10.2

| Критерии оценивания результатов обучения | Планируемые результаты обучения (показатели достижения заданного уровня освоения компетенций) | Уровень освоения компетенции | | | |
|--|---|--|---|--|---|
| | | Не сформирован (0 баллов) | Пороговый уровень (3 балла) | Базовый уровень (4 балла) | Продвинутый уровень (5 баллов) |
| 1 | 2 | 3 | 4 | 5 | 6 |
| Полнота знаний | ОПК 3.1 | Уровень знаний ниже минимальных требований. Имеют место грубые ошибки. | Демонстрирует общие знания базовых понятий по темам/разделам дисциплины. Допускается значительное количество негрубых ошибок. | Уровень знаний соответствует программе подготовки по темам/разделам дисциплины. Допускается несколько негрубых/несущественных ошибок. Не отвечает на дополнительные вопросы. | Уровень знаний соответствует программе подготовки по темам/разделам дисциплины. Свободно и аргументированно отвечает на дополнительные вопросы. |
| Наличие умений | ОПК 3.2 ОПК 3.3 | Отсутствие минимальных умений. Не умеет решать | Продемонстрированы частично основные умения. | Продемонстрированы все основные умения. Решены все основные задания с | Продемонстрированы все основные умения. Решены все основные |

| | | | | | |
|--|--|--|--|--------------------------------------|---|
| | | стандартные задачи. Имеют место грубые ошибки. | Решены типовые задачи. Допущены негрубые ошибки. | негрубыми ошибками или с недочетами. | задания в полном объеме без недочетов и ошибок. |
|--|--|--|--|--------------------------------------|---|

10.3. Типовые контрольные задания и материалы, необходимые для оценки результатов обучения

Перечень типовых заданий для проведения текущего контроля успеваемости обучающихся.

1. Написать скрипт для построения ROC-кривой, PR-кривой, нахождения площади под обеими кривыми для произвольного двоичного классификатора.
2. Реализуйте метод потенциальных функций на примере набора данных Iris.
3. Используйте библиотеку sklearn для классификации набора данных MNIST с классификаторами SVC, метод ближайших соседей, наивный байесовский классификатор, решающие деревья. Покажите преимущества одних по сравнению с другими
4. Воспользовавшись готовыми библиотеками (xgboost, catboost, lightgbm), постройте регрессор для цены на домовые в участке в Америке.
5. Скачав данные с <https://www.kaggle.com/c/pubg-finish-placement-prediction> постройте алгоритм, предсказывающий место игрока в онлайн игре.

Типовые вопросы для экзамена

Билет 1

1. Решающие деревья.
2. Алгоритм DBSCAN.

Билет 2

1. Градиентный бустинг.
2. Логистическая регрессия.

Билет 3

1. Метод наименьших квадратов.
2. Нейронные сети.

Билет 4

1. Метод k-ближайших соседей.
2. Решающий лес.

Билет 5

1. Разложение ошибки на разброс и смещение.
2. Метод опорных векторов.

Билет 6

1. Микро и макроусреднение.
2. Бэггинг по случайным подпространствам.

Билет 7

1. Точность и полнота.
2. Формула Надарая-Ватсона.

Билет 8

1. Обучение с учителем/без учителя.

2. Кластеризация методом к-средних.

Билет 9

1. xgboost, catboost, lightgbm — основные идеи и особенности реализации.
2. Метод главных компонент.

Билет 10

1. Наивный байесовский классификатор.
2. Метод опорных векторов с нелинейной разделяющей поверхностью.

Оценочные материалы по промежуточной аттестации, предназначенные для проверки соответствия уровня подготовки по дисциплине требованиям СУОС, хранятся на кафедре-разработчике РПД в печатном и электронном виде.

**Лист актуализации рабочей программы
по дисциплине «Введение в машинное обучение»
Направление: 03.03.02 Физика
Направленность (профиль): Физическая информатика**

| № | Характеристика внесенных изменений (с указанием пунктов документа) | Дата и № протокола Учёного совета ФФ НГУ | Подпись ответственного |
|---|--|--|------------------------|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |