

**Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Московский физико-технический институт  
(национальный исследовательский университет)»**

**УТВЕРЖДЕНО**  
**Директор физтех-школы**  
**прикладной математики**  
**и информатики**

**А.М. Райгородский**

**Рабочая программа дисциплины (модуля)**

<b>Дисциплина:</b>	Основные методы кластеризации и распознавания
<b>Направление:</b>	Прикладные математика и физика
<b>Профиль подготовки:</b>	Математическая физика, компьютерные технологии и математическое моделирование в экономике Физтех-школа прикладной математики и информатики Кафедра проблем передачи информации и анализа данных
<b>Курс:</b>	4
<b>Квалификация:</b>	Бакалавр

Семестр, формы промежуточной аттестации: 7 (Осенний) – Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

практические (семинарские) занятия: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час. всего, в том числе:

задания, курсовые работы: 0 час.

Подготовка к экзамену: 0 час.

Всего часов: 45, всего зач.ед.: 1

**Программу составили:** **С.М. Карпенко, преподаватель**  
**М.Ю. Гончаров, преподаватель**

**Аннотация**

Курс направлен на вооружение студентов сведениями и компетенциями, полезными для успешной работы в области машинного обучения. Мы стараемся сочетать углубленный теоретический анализ стандартных задач и алгоритмов с обзорными сведениями об их актуальном использовании в современной практике. В качестве исходных постановок мы рассматриваем бинарную классификацию и кластеризацию с известным числом классов. Простейшие алгоритмы (logistic regression, MLP, k-means) для решения этих задач студентам уже из-

вестны. Для углубленного анализа этих алгоритмов вводим некоторые понятия статистической теории обучения (VC-размерность, double-descent). Отталкиваясь от этих понятий, мы получаем средства к критическому анализу алгоритмов для более современных и реалистичных постановок.

## **Содержание дисциплины (модуля), структурированное по темам (разделам)**

Семестр: 7 (Осенний)

### 1. Введение.

Обучаемые (параметризованные) алгоритмы. Приложения. Данные, признаки. Обзор протоколов обучения: с учителем, без учителя, с подкреплением. Примеры. Роль методов оптимизации. Нейронные сети, коннекционизм. Переобучение и регуляризация.

### 2. Обучение с учителем.

Логистическая регрессия, перцептрон. Обратное распространение ошибки. Многослойный перцептрон. Разделяемые веса. Обучаемые метрики (сиамские сети). Конволютивные сети. Машины опорных векторов (SVM). Квадратичная оптимизация. Нестандартные скалярные произведения.

Простейший алгоритм обучения, использующий теорему Байеса.

Практические вопросы: сбор базы данных, выбор признаков, диагностика качества работы алгоритма. Типы ошибок, характеристическая кривая (ROC-curve).

Деревья принятия решений, Алгоритм C4.5. Бустинг. Алгоритм Виолы-Джонса.

### 3. Обучение без учителя и анализ данных.

Кластеризация. Иерархическая кластеризация. Алгоритм К-средних. Модель смеси гауссиан. Дискриминант Фишера. Алгоритм ожидания-максимизации (EM).

Сокращение размерности. Анализ главных компонент (PCA). Нейронные сети, осуществляющие нелинейный анализ главных компонент.

Пропущенные данные. Вероятностная трактовка PCA.

Факторный анализ, анализ независимых компонент.

### 4. Вероятностный вывод.

Скрытые Марковские модели. Марковские случайные поля. Общий взгляд на модели со скрытыми параметрами.

## **Перечень основной и дополнительной литературы, необходимой для освоения дисциплины (модуля)**

Основная литература

1. Шалев-Шварц Ш., Бен-Давид Ш. Идеи машинного обучения / ДМК-Пресс, 2019.
2. Бенджио И., Гудфеллоу Я., Курвилль А. Глубокое обучение / ДМК-Пресс, 2018.
3. Николенко С., Архангельская Е., Кадурын А. Глубокое обучение - погружение в мир нейронных сетей / Питер, 2018.
4. Мерков А.Б. Введение в методы статистического обучения // Материалы к учебнику: <http://www.recognition.mccme.ru/pub/RecognitionLab.html/sltb.pdf>.

5. Bishop C. Pattern Recognition and Machine Learning. Springer, 2006. - 101 p.
6. Ветров Д.П., Кропотов Д.А. Алгоритмы выбора моделей и построения коллективных решений в задачах классификации, основанные на принципе устойчивости. М.: КомКнига, 2006. -112 с.
7. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. - 511 с.

Дополнительная литература

1. Вьюгин В.В. Элементы математической теории машинного обучения // Учебное пособие. М.: МФТИ-ИППИ РАН, 2012. - 323 с.

**Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

<https://arxiv.org/>

<https://openreview.net/>

<https://paperswithcode.com/>

<https://github.com/>