

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

**УТВЕРЖДЕНО**  
**Проректор по учебной работе и**  
**довузовской подготовке**  
**А.А. Воронов**

**Рабочая программа дисциплины (модуля)**

<b>по дисциплине:</b>	Основные методы кластеризации и распознавания
<b>по направлению:</b>	Прикладная математика и физика
<b>профиль подготовки:</b>	Компьютерные технологии и интеллектуальный анализ данных 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 зачёт

**Программу составил:** **С.М. Карпенко, 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 известны. 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.

В этом году в обзорной части обсуждались такие алгоритмы, как HDBSCAN(кластеризация), UMAP(снижение размерности), GAN/VAE, transformers, TRPO.

**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, персептрон. Обратное распространение ошибки. Многослойный персептрон. 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 рекомендуемой литературы

### Основная литература

1. Мерков А.Б. Введение в методы статистического обучения // Материалы к учебнику: <http://www.recognition.mccme.ru/pub/RecognitionLab.html/sltb.pdf>.

2. Bishop C. Pattern Recognition and Machine Learning. Springer, 2006. - 101 p.

3. 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.

4. 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.