

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

УТВЕРЖДЕНО

**Директор института нано-, био-,
информационных, когнитивных
и социогуманитарных наук и
технологий**

Т.Е. Григорьев

Рабочая программа дисциплины (модуля)

по дисциплине:	Методы машинного обучения
по направлению:	Прикладные математика и физика
профиль подготовки:	Термоядерная энергетика и плазменные технологии Физтех-школа природоподобных, плазменных и ядерных технологий им. И.В. Курчатова кафедра информатики и вычислительных сетей
курс:	4
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 7 (осенний) - Дифференцированный зачет

Аудиторных часов: 45 всего, в том числе:

лекции: 0 час.

семинары: 0 час.

лабораторные занятия: 45 час.

Самостоятельная работа: 45 час.

Всего часов: 90, всего зач. ед.: 2

Количество контрольных работ, заданий: 1

Программу составил: А.Б. Теслюк, старший преподаватель

Программа обсуждена на заседании кафедры информатики и вычислительных сетей 18.03.2021

Аннотация

В рамках курса изучаются классические методы машинного обучения такие как метод максимального правдоподобия, линейная регрессия, нейронные сети, решающие деревья и другие,

Раздел машинного обучения, с одной стороны, образовался в результате разделения науки о нейросетях на методы обучения сетей и виды топологий их архитектуры, с другой стороны — вобрал в себя методы математической статистики. Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения. Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам. Многие методы тесно связаны с извлечением информации и интеллектуальным анализом данных (Data Mining).

1. Цели и задачи

Цель дисциплины

- знакомство студентов с основополагающими подходами, которые применяются для поиска функциональных закономерностей из эмпирических данных;
- знакомство студентов с широким спектром инструментов для решения задач поиска закономерностей в данных.

Задачи дисциплины

- знакомство с основными понятиями, применяемыми в машинном обучении. Постановка задач машинного обучения;
- обзор методов, применяемых в задачах кластеризации, классификации и регрессии;
- получение навыков применения методов машинного обучения для решения задач.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные методы и алгоритмы, применяемые при поиске закономерностей в эмпирических данных;
- подходы для организации управления большими массивами данных.

уметь:

- применять методы и алгоритмы для решения прикладных задач анализа данных;
- реализовывать алгоритмы в виде компьютерных программ.

владеть:

- специальной терминологией в области машинного обучения;
- методологией и навыками решения научных и практических задач с использованием методов машинного обучения.

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

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Основные понятия.			5	10
2	Основные методы.			10	5
3	Регрессии.			5	10
4	Методы кластеризации.			10	5
5	Решающие деревья и композиции алгоритмов.			5	10
6	Нейронные сети.			10	5
Итого часов				45	45
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 7 (Осенний)

1. Основные понятия.

Основные понятия: примеры постановки задач машинного обучения, виды признаков, типы задач, метрики.

2. Основные методы.

Метод максимального правдоподобия. Понятие функционала правдоподобия в случае зависимых и независимых наблюдений. Примеры построения оценок максимального правдоподобия для параметров ряда различных распределений.

Байесовский подход. Понятие условных вероятностей, апостериорного распределения. Формула Байеса. Принцип максимума апостериорной вероятности.

Метод стохастического градиента.

Метод опорных векторов. Понятие зазора между классами, оптимальной разделяющей гиперплоскости, опорных векторов. VC-размерность. Функции ядра, спрямляющее пространство.

3. Регрессии.

Многомерная линейная регрессия. Задача регрессии. Построение оценок максимального правдоподобия для параметров многомерной линейной регрессии.

Логистическая регрессия. Сигмовидная функция, принцип максимума правдоподобия, логарифмическая функция потерь.

4. Методы кластеризации.

Методы кластеризации. Задачи кластеризации, основные типы алгоритмов кластеризации. Иерархическая кластеризация, EM-алгоритм, метод k-средних, карты Кохонена.

5. Решающие деревья и композиции алгоритмов.

Решающие деревья и композиции алгоритмов. Решающий список, решающее дерево, Алгоритмы ID3, CART. Варианты линейных и стохастических композиций алгоритмов.

6. Нейронные сети.

Нейронные сети. Модель перцептрона, функции активации, вопросы полноты, теорема Колмогорова. Алгоритм обратного распространения ошибок.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная мультимедиа проектором и экраном, маркерной доской.

Локальная сеть, связывающая рабочие места с сервером, выход в интернет.

6. Перечень рекомендуемой литературы

Основная литература

Фонд литературы кафедры

1. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979
2. Мерков А. Б. Распознавание образов. Введение в методы статистического обучения. 2011

Дополнительная литература

Фонд литературы кафедры

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014

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

Handbook of Applied Cryptography, by A.Menezes, P. van Oorschot, and S. Vanstone, CRC Press, 1996. For further information, www.cacr.math.uwaterloo.ca/hac

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

на лабораторных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Для успешного освоения курса, помимо посещения лабораторных занятий, от студентов требуется самостоятельная работа в объеме не менее чем те часы, которые указаны для каждого раздела программы (всего не менее 45 часов в семестр). В основном, это время отводится на проработку материала лабораторных, а также на подготовку к промежуточным тестированиям, которые проводятся для текущего контроля за усвоением материала. Студенты, успешно прошедшие все формы промежуточного контроля, допускаются к сдаче дифференцированного зачета по дисциплине.

ПРИЛОЖЕНИЕ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Термоядерная энергетика и плазменные технологии Физтех-школа природоподобных, плазменных и ядерных технологий им. И.В. Курчатова кафедра информатики и вычислительных сетей
курс:	<u>4</u>
квалификация:	бакалавр
Семестр, формы промежуточной аттестации: 7 (осенний) - Дифференцированный зачет	
Разработчик:	А.Б. Теслюк, старший преподаватель

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов

2. Показатели оценивания компетенций

В результате изучения дисциплины «Методы машинного обучения» обучающийся должен:

знать:

- основные методы и алгоритмы, применяемые при поиске закономерностей в эмпирических данных;
- подходы для организации управления большими массивами данных.

уметь:

- применять методы и алгоритмы для решения прикладных задач анализа данных;
- реализовывать алгоритмы в виде компьютерных программ.

владеть:

- специальной терминологией в области машинного обучения;
- методологией и навыками решения научных и практических задач с использованием методов машинного обучения.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

В целях текущего контроля успеваемости предусмотрен краткий опрос по темам предыдущих занятий по теме прошлой лекции или в конце занятия по пройденной теме.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Общая формула байесовского классификатора.
2. Метод стохастического градиента.
3. Вероятностный смысл регуляризации. Типы регуляризаторов.
4. Постановка задачи многомерной линейной регрессии.
5. Метод главных компонент.
6. Метод обратного распространения ошибок.
7. Определение алгоритмической композиции.
8. Достоинства и недостатки алгоритма AdaBoost.
9. Основные цели кластеризации.

Критерии оценивания

отлично

10 Выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

9 Выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

8 Выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

хорошо

7 Выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

6 Выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

5 Выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

удовлетворительно

4 Выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

3 Выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

неудовлетворительно

2 Выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

1 Выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

При проведении дифференцированного зачета обучающемуся предоставляется не менее 40 минут на подготовку. Опрос и ответы на дополнительные вопросы не должны превышать 0,5 астрономических часа. По завершении отведенного на опрос времени, экзаменатор должен выставить обучающемуся оценку. Оценка включает в себя кроме результатов опроса, текущую успеваемость.