

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

УТВЕРЖДЕНО
Директор

А.В. Малеев

	Рабочая программа дисциплины (модуля)
по дисциплине:	Дополнительные главы машинного обучения
по направлению:	Программная инженерия
профиль подготовки:	Разработка программно-информационных систем высшая школа программной инженерии высшая школа программной инженерии МФТИ - Яндекс
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 48 час.

Всего часов: 108, всего зач. ед.: 3

Программу составил: А.В. Малеев, директор

Программа обсуждена на заседании высшей школы программной инженерии МФТИ - Яндекс 08.06.2022

Аннотация

Данная дисциплина разработана для обучения студентов основным теориям, концепциям и инструментам машинного обучения, а также формирования у студентов прикладных навыков в области machine learning. Основная задача данной дисциплины - сформировать у обучающихся навыки практического решения задач анализа данных и машинного обучения.

1. Цели и задачи

Цель дисциплины

- обучение студентов основным теориям, концепциям и инструментам машинного обучения, а также формирование у студентов прикладных навыков в области machine learning.

Задачи дисциплины

- обучить правильно формулировать задачу в терминах машинного обучения;
- сформировать у обучающихся навыки практического решения задач анализа данных и машинного обучения.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде	УК-3.1 Способен устанавливать разные виды коммуникации (учебную, научную, деловую, неформальную и др.)
	УК-3.2 Взаимодействует с другими членами команды для достижения поставленной задачи
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.2 Способен планировать самостоятельную деятельность в решении профессиональных задач; подвергать критическому анализу проделанную работу; находить и творчески использовать имеющийся опыт в соответствии с задачами саморазвития
ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК-2.2 Владеет навыками создания программного обеспечения для ЭВМ и систем различной архитектуры
ОПК-3 Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	ОПК-3.1 Обладает навыками разработки архитектуры программных систем и компонентов с учетом требований к производительности, надежности и безопасности
	ОПК-3.3 Знает основы информационной безопасности и методы защиты программного обеспечения от угроз и атак
ОПК-8 Способен осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий	ОПК-8.2 Умеет применять технологии машинного обучения в различных прикладных областях
ПК-1 Способен самостоятельно или в качестве члена малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-1.2 Способен проводить научные исследования самостоятельно или в качестве члена малого научного коллектива
ПК-2 Способен формализовать и алгоритмизировать поставленную задачу	ПК-2.1 Способен формализовать и алгоритмизировать поставленную задачу

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные принципы и способы постановки задач машинного обучения;
- принципы и проблематику теории обучения машин, основные современные методы обучения по прецедентам;
- методы оценивания и комбинирования по классификаторам.

уметь:

- формализовать постановки прикладных задач анализа данных;
- использовать современные методы обучения по прецедентам для решения задач машинного обучения;
- оценивать точность и эффективность полученных решений.

владеть:

- основными понятиями теории машинного обучения;
- современным математическим аппаратом прикладного анализа данных.

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	Композиции классификаторов, бустинг	2	2		4
2	Критерии выбора моделей	4	4		4
3	Методы отбора признаков	4	4		6
4	Методы ранжирования	4	4		4
5	Обучение с подкреплением	4	4		6
6	Задачи с частичным обучением	4	4		4
7	Оценивание и комбинирование классификаторов	2	2		4
8	Тематическое моделирование	2	2		6
9	Байесовское обучение	2	2		4
10	Введение в глубинное обучение	2	2		6
Итого часов		30	30		48
Подготовка к экзамену		0 час.			
Общая трудоёмкость		108 час., 3 зач.ед.			

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

Семестр: 6 (Весенний)

1. Композиции классификаторов, бустинг

Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция.

Взвешенное голосование.

Алгоритм AdaBoost. Экспоненциальная аппроксимация пороговой функции потерь. Процесс последовательного обучения базовых алгоритмов. Теорема о сходимости бустинга.

Обобщение бустинга как процесса градиентного спуска. Теорема сходимости. Алгоритм AnyBoost.

Простое голосование (комитет большинства). Эвристический алгоритм ComBoost. Идентификация нетипичных объектов (выбросов). Обобщение на большое число классов.

Решающий список (комитет старшинства). Эвристический алгоритм. Стратегия выбора классов для базовых алгоритмов.

Выпуклые функции потерь. Методы построения смесей: последовательный и иерархический. Построение смесей экспертов с помощью EM-алгоритма.

2. Критерии выбора моделей

Внутренние и внешние критерии.

Эмпирические и аналитические оценки функционала полного скользящего контроля.

Скользящий контроль, разновидности эмпирических оценок скользящего контроля.

Критерий непротиворечивости.

Регуляризация. Критерий Акаике (AIC). Байесовский информационный критерий (BIC).

Агрегированные и многоступенчатые критерии.

3. Методы отбора признаков

Усечённый поиск в ширину, многорядный итерационный алгоритм МГУА.

Генетический алгоритм, его сходство с МГУА.

Случайный поиск и Случайный поиск с адаптацией (СПА).

4. Методы ранжирования

Постановка задачи ранжирования.

Примеры прикладных задач.

Признаки в задаче ранжирования поисковой выдачи: текстовые, ссылочные, кликовые.

Критерии качества ранжирования.

Точечный, попарный и списочный подходы.

5. Обучение с подкреплением

Задача о многоруком бандите. Жадные и эпсилон-жадные стратегии. Среда для экспериментов. Метод сравнения с подкреплением. Метод преследования.

Адаптивные стратегии на основе скользящих средних.

Уравнения Беллмана. Оптимальные стратегии. Динамическое программирование. Метод итераций по ценностям и по стратегиям.

Методы временных разностей: TD, SARSA, Q-метод. Многошаговое TD-прогнозирование.

Адаптивный полужадный метод VDBE.

6. Задачи с частичным обучением

Постановка задачи Semisupervised Learning, примеры приложений.

Простые эвристические методы: self-training, co-training, co-learning.

Адаптация алгоритмов кластеризации для решения задач с частичным обучением.

Кратчайший незамкнутый путь. Алгоритм Ланса-Уильямса. Алгоритм k-средних.

Трансдуктивный метод опорных векторов TSVM.

Алгоритм Expectation-Regularization на основе многоклассовой регуляризированной логистической регрессии.

7. Оценивание и комбинирование классификаторов

Минимизация средних потерь типа 0-1.
Комбинирование классификаторов.
Баггинг.
Алгоритм ада-бустинга.

8. Тематическое моделирование

Задачи тематического моделирования, коллекции текстовых документов и матрица документы—слова. Перплексия как мера качества тематической модели. Задача тематического поиска.

Униграммная модель документа. Метод максимума правдоподобия и метод максимума апостериорной вероятности. Применение метода множителей Лагранжа.

Вероятностный латентный семантический анализ PLSA. ЕМ-алгоритм. Инкрементное добавление новых документов (folding-in). Задача с частичным обучением.

Робастная тематическая модель с фоновой и шумовой компонентой. Эксперименты по сравнению робастных и регуляризованных моделей.

9. Байесовское обучение

Понятие условной независимости, графические модели.

Байесовские сети.

Марковские поля.

Скрытые марковские модели.

Условные случайные поля.

10. Введение в глубинное обучение

Рекуррентные нейросети, сверточные нейросети.

Примеры прикладных задач, успешно решаемых с помощью глубинного обучения.

Ограниченная машина Больцмана.

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

компьютер, мультимедийное оборудование (звуковая система, проектор).

6. Перечень рекомендуемой литературы

Основная литература

1. Машинное обучение [Текст]/Х. Бринк, Дж. Ричардс, М. Феверолф, Real-World Machine Learning, -СПб., Питер, 2017
2. Математические основы машинного обучения и прогнозирования, Электронная версия печатной публикации / В. В. Вьюгин. — Москва, МЦНМО, 2014

-

Дополнительная литература

Литература, рекомендованная для самостоятельного изучения:

- Айвазян С. А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Классификация и снижение размерности. — М. Финансы и статистика. 1989.
- Хардле В. Прикладная непараметрическая регрессия. — М.: Мир. 1993.

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

Не используются

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

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

В процессе самостоятельной работы обучающихся предполагается использование таких программных средств, как WEKA, IPython Notebook и др.

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

Программа курса в разделе "самостоятельная работа" приводит минимальное необходимое количество часов для освоения темы. Самостоятельная работа включает в себя:

- проработку учебного материала (по конспектам лекций, учебной и научной литературе. Конспектировать материал лекций является желательным, но не обязательным условием для успешного освоения дисциплины);
- подготовку к практическим занятиям, выполнение домашних теоретических и практических заданий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Программная инженерия
профиль подготовки:	Разработка программно-информационных систем высшая школа программной инженерии МФТИ - Яндекс высшая школа программной инженерии
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Разработчик: А.В. Малеев, директор

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде	УК-3.1 Способен устанавливать разные виды коммуникации (учебную, научную, деловую, неформальную и др.)
	УК-3.2 Взаимодействует с другими членами команды для достижения поставленной задачи
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.2 Способен планировать самостоятельную деятельность в решении профессиональных задач; подвергать критическому анализу проделанную работу; находить и творчески использовать имеющийся опыт в соответствии с задачами саморазвития
ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК-2.2 Владеет навыками создания программного обеспечения для ЭВМ и систем различной архитектуры
ОПК-3 Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	ОПК-3.1 Обладает навыками разработки архитектуры программных систем и компонентов с учетом требований к производительности, надежности и безопасности
	ОПК-3.3 Знает основы информационной безопасности и методы защиты программного обеспечения от угроз и атак
ОПК-8 Способен осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий	ОПК-8.2 Умеет применять технологии машинного обучения в различных прикладных областях
ПК-1 Способен самостоятельно или в качестве члена малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-1.2 Способен проводить научные исследования самостоятельно или в качестве члена малого научного коллектива
ПК-2 Способен формализовать и алгоритмизировать поставленную задачу	ПК-2.1 Способен формализовать и алгоритмизировать поставленную задачу

2. Показатели оценивания компетенций

В результате изучения дисциплины «Дополнительные главы машинного обучения» обучающийся должен:

знать:

- основные принципы и способы постановки задач машинного обучения;
- принципы и проблематику теории обучения машин, основные современные методы обучения по прецедентам;
- методы оценивания и комбинирования по классификаторам.

уметь:

- формализовать постановки прикладных задач анализа данных;
- использовать современные методы обучения по прецедентам для решения задач машинного обучения;
- оценивать точность и эффективность полученных решений.

владеть:

- основными понятиями теории машинного обучения;
- современным математическим аппаратом прикладного анализа данных.

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

С целью контроля освоения обучающимися учебного материала проводится устный опрос в начале занятия по теме прошлого занятия.

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

Примерные вопросы для проведения промежуточной аттестации обучающихся:

1. Опишите метод множителей Лагранжа.
2. Что представляет из себя вероятностный семантический анализ PLSA? Приведите примеры.
3. Опишите эксперименты по сравнению робастных и регуляризованных моделей.
4. В чем заключается отличие рекуррентных нейронных сетей от сверточных?
5. Опишите эмпирические и аналитические оценки функционала полного скользящего контроля.
6. Что представляет собой безусловная задача оптимизации?
7. Приведите пример задачи, для решения которой можно использовать метод множителей Лагранжа
8. Дайте пример прикладных задач, успешно решаемых с помощью глубинного обучения.
9. Опишите особенности рекуррентной нейронной сети.
10. Дайте определение баггинга

Критерии оценивания

- оценка «отлично (10)» 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
- оценка «отлично (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений
- оценка «хорошо (7)» 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;
- оценка «хорошо (6)» выставляется студенту, если он знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (5)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «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 владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;

- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины

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

Дифференцированный зачет проводится по итогам текущей успеваемости и сдачи заданий и других видов работ, предусмотренных программой дисциплины и (или) путем организации специального опроса, проводимого в устной и (или) письменной форме.