

**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)
по дисциплине:	Онлайн-методы в машинном обучении
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра математических основ управления
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 45 всего, в том числе:

лекции: 30 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Всего часов: 90, всего зач. ед.: 2

Количество контрольных работ, заданий: 4

Программу составил: Н.А. Пучкин, канд. физ.-мат. наук

Программа обсуждена на заседании кафедры математических основ управления 03.04.2024

Аннотация

В курсе рассматриваются ключевые понятия и современные математические методы теории онлайн-машинного обучения, а также их приложения для прогнозирования в различных областях науки. Характерная черта данного подхода — отсутствие предположения о существовании стохастических механизмов, генерирующих данные. Вместо этого рассматриваются конкурирующие классы методов прогнозирования, на основании которых строится оптимальный метод, который прогнозирует не хуже, чем любой метод из заданного класса. Первая часть курса содержит обзор основных понятий и классических алгоритмов Литлстоуна и Вармута, Вовка, Фройнда и Шапире. Получены оценки регрета этих алгоритмов. Приведены также их приложения для построения алгоритма бустинга AdaBoost и универсального портфеля (финансовых инструментов) Ковера. Представлена теория предсказания в условиях неполной информации.

В второй части излагается теория онлайн-выпуклой оптимизации - алгоритмы следования за регуляризованным лидером, алгоритмы онлайн-градиентного спуска и онлайн-зеркального спуска. В третьей части теория предсказания индивидуальных последовательностей исходов рассматривается в свете теории игр. Доказана минимаксная теорема и изучаются различные виды равновесия в играх с полной информацией. Приводятся алгоритмы построения калибруемых по Дейvidу предсказаний. Доказана теорема Блекуэла о достижимости и рассмотрены ее приложения.

Для успешного освоения курса слушателю желательно знать основные разделы общей теории алгоритмов и обладать элементарными знаниями в области теории вероятностей. Слушателям курса предлагается большое количество задач по указанным темам.

1. Цели и задачи

Цель дисциплины

- изучение основных понятий и методов онлайн-методов в машинном обучении.

Задачи дисциплины

- освоение студентами базовых знаний в области машинного обучения;
- приобретение теоретических знаний в области байесовской теории машинного обучения;
- оказание консультаций и помощи студентам в решении теоретических и практических задач.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1 Формулирует в рамках обозначенной проблемы, цель, задачи, актуальность, значимость (научную, практическую, методическую и иную в зависимости от типа проекта), ожидаемые результаты и возможные сферы их применения
	УК-2.2 Способен прогнозировать результат деятельности и планировать последовательность шагов для достижения данного результата. Формирует план-график реализации проекта в целом и план контроля его выполнения
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии	ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения
	ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности
	ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий

ПК-2.4 Владеет методами и алгоритмами решения задач цифровой обработки сигналов, использования сети Интернет, аннотирования, реферирования, библиографического поиска, опыт работы с научными источниками	
ПК-3 Владеет навыками участия в научных дискуссиях, выступления с сообщениями и докладами устного, письменного и виртуального (размещение в информационных сетях) характера, представления материалов собственных исследований	ПК-3.1 Знает основы ведения научной дискуссии и формы устного научного высказывания
	ПК-3.2 Умеет вести корректную дискуссию в области информационных технологий задавать вопросы и отвечать на поставленные вопросы по теме научной работы
	ПК-3.3 Имеет практический опыт участия в научных студенческих конференциях, очных, виртуальных, заочных обсуждениях научных проблем в области информационных технологий

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- фундаментальные понятия, законы, методы статистической теории машинного обучения;
- основные свойства соответствующих математических объектов;
- аналитические и численные подходы и методы для решения типовых прикладных задач.

уметь:

- понять поставленную задачу;
- использовать свои знания для решения фундаментальных и прикладных задач теории машинного обучения;
- оценивать корректность постановок задач;
- строго доказывать или опровергать утверждение;
- самостоятельно находить алгоритмы решения задач теории машинного обучения, в том числе и нестандартных, и проводить их анализ;
- самостоятельно видеть следствия полученных результатов.

владеть:

- навыками освоения большого объема информации и решения задач;
- навыками самостоятельной работы и освоения новых дисциплин;
- культурой постановки, анализа и решения математических и прикладных задач, требующих для своего решения использования математических подходов и методов теории машинного обучения;
- предметным языком дискретной математики и навыками грамотного описания решения задач и представления полученных результатов.

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	Постановка задачи предсказания с использованием экспертных стратегий.	10	5		15
2	Алгоритм экспоненциального взвешивания экспертных решений.	10	5		15
3	Применение различных алгоритмов для многомерной регрессии.	10	5		15

Итого часов	30	15		45
Подготовка к экзамену	0 час.			
Общая трудоёмкость	90 час., 2 зач.ед.			

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

Семестр: 2 (Весенний)

1. Постановка задачи предсказания с использованием экспертных стратегий.

Постановка задачи предсказания с использованием экспертных стратегий. Понятие (внешнего) регрета. Алгоритм взвешенного большинства. Оценка числа ошибок.

Алгоритм оптимального распределения потерь в режиме онлайн. Оценка его регрета.

Алгоритм следования за возмущенным лидером. Оценка его регрета. Состоятельность по Ханнану.

Задача минимизации регрета с точки зрения теории оптимизации. Алгоритм следования за регуляризованным лидером.

Онлайн-метод градиентного спуска.

Онлайн-метод зеркального спуска.

Неравенства больших уклонений. Варианты неравенства Хефдинга.

2. Алгоритм экспоненциального взвешивания экспертных решений.

Алгоритм экспоненциального взвешивания экспертных решений.

Алгоритм экспоненциального взвешивания экспертных решений с переменным параметром обучения.

Задача о многоруком бандите. Стохастическая и детерминированная постановки. Алгоритм для детерминированной постановки.

Бустинг. Алгоритм Адабуст и его свойства.

Агрегирующий алгоритм Вовка. Конечное и бесконечное множество экспертов.

3. Применение различных алгоритмов для многомерной регрессии.

Применение агрегирующего алгоритма для различных функций потерь: логарифмической, квадратичной, простой.

Универсальный портфель акций Ковера. Алгоритм построения портфеля. Оценка

Применение агрегирующего алгоритма для многомерной регрессии в режиме онлайн.

Калибруемость предсказаний по Дэвиду. Алгоритмы построения калибруемых предсказаний.

Универсальные RKHS. Построение универсальных алгоритмов для онлайн регрессии на основе метода калибруемости.

Средние Радемахера и оценка калибровочной ошибки.

Агрегирующий алгоритм как результат построения калибруемых предсказаний.

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

Необходимое оборудование для лекций и практических занятий: компьютер и мультимедийное оборудование (проектор, звуковая система).

6.Перечень рекомендуемой литературы

Основная литература

1. Колмогоровская сложность и алгоритмическая случайность [Текст] : учеб. пособие для вузов / В. В. Вьюгин ; М-во образования и науки РФ, Моск. физ.-техн. ин-т (гос. ун-т), Ин-т проблем информации им. А. А. Харкевича .— М. : МФТИ, 2012 .— 140 с.

Дополнительная литература

1. Теория игр [Текст] : учеб. пособие для вузов : рек. М-вом общ. и проф. образования РФ / Л. А. Петросян, Н. А. Зенкевич, Е. А. Семина .— М. : Высшая школа, 1998 .— 304 с.

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

<http://www.iitp.ru/ru/userpages/>

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

Не предусмотрено.

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

Студент, изучающий курс "Онлайн-методы в машинном обучении", должен, с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике.

В результате изучения дисциплины студент должен знать основные определения, понятия, аксиомы, методы доказательств.

Успешное освоение курса требует напряжённой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы,
- проработку учебного материала (по конспектам лекций, учебной и научной литературе), подготовку ответов на вопросы, предназначенных для самостоятельного изучения, доказательство отдельных утверждений, свойств;
- решение задач, предлагаемых студентам на практических занятиях и в качестве курсового задания,
- подготовку к дифференцированному зачету.

Руководство и контроль за самостоятельной работой студента осуществляется в форме индивидуальных консультаций.

Показателем владения материалом служит умение решать задачи. Для формирования умения применять теоретические знания на практике студенту необходимо решать как можно больше задач. При решении задач каждое действие необходимо аргументировать, ссылаясь на известные теоретические сведения.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к лектору или преподавателю, ведущему практические занятия.

Литература для самостоятельной работы:

1. Bousquet, O., Boucheron, S., and Lugosi, G.: Introduction to statistical learning theory. in: Advanced Lectures on Machine Learning. pp. 169--207 (2004)
2. Steinwart, I.: On the influence of the kernel on the consistency of support vector machines. Journal of Machine Learning Research. 2, 67--93 (2001)
3. Rakhlin, A., Sridharan, K., Tewari, A.: Online learning: Beyond regret. In Proceedings of the 24rd Annual Conference on Learning Theory, v/ 19 of JMLR Workshop and Conference Proceedings, pages 559--594, 2011. longer version available as arXiv:1011.3168 (2011).

4. S Bubeck and N. Cesa-Bianchi, Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. In Foundations and Trends in Machine Learning, Vol 5: No 1, 1-122, 2012.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра математических основ управления
курс:	<u>1</u>
квалификация:	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Разработчик: Н.А. Пучкин, канд. физ.-мат. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1 Формулирует в рамках обозначенной проблемы, цель, задачи, актуальность, значимость (научную, практическую, методическую и иную в зависимости от типа проекта), ожидаемые результаты и возможные сферы их применения
	УК-2.2 Способен прогнозировать результат деятельности и планировать последовательность шагов для достижения данного результата. Формирует план-график реализации проекта в целом и план контроля его выполнения
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии	ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения
	ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности
	ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий
	ПК-2.4 Владеет методами и алгоритмами решения задач цифровой обработки сигналов, использования сети Интернет, аннотирования, реферирования, библиографического поиска, опыт работы с научными источниками
ПК-3 Владеет навыками участия в научных дискуссиях, выступления с сообщениями и докладами устного, письменного и виртуального (размещение в информационных сетях) характера, представления материалов собственных исследований	ПК-3.1 Знает основы ведения научной дискуссии и формы устного научного высказывания
	ПК-3.2 Умеет вести корректную дискуссию в области информационных технологий задавать вопросы и отвечать на поставленные вопросы по теме научной работы
	ПК-3.3 Имеет практический опыт участия в научных студенческих конференциях, очных, виртуальных, заочных обсуждениях научных проблем в области информационных технологий

2. Показатели оценивания компетенций

В результате изучения дисциплины «Онлайн-методы в машинном обучении» обучающийся должен:

знать:

- фундаментальные понятия, законы, методы статистической теории машинного обучения;
- основные свойства соответствующих математических объектов;
- аналитические и численные подходы и методы для решения типовых прикладных задач.

уметь:

- понять поставленную задачу;
- использовать свои знания для решения фундаментальных и прикладных задач теории машинного обучения;
- оценивать корректность постановок задач;
- строго доказывать или опровергать утверждение;
- самостоятельно находить алгоритмы решения задач теории машинного обучения, в том числе и нестандартных, и проводить их анализ;
- самостоятельно видеть следствия полученных результатов.

владеть:

- навыками освоения большого объема информации и решения задач;
- навыками самостоятельной работы и освоения новых дисциплин;
- культурой постановки, анализа и решения математических и прикладных задач, требующих для своего решения использования математических подходов и методов теории машинного обучения;
- предметным языком дискретной математики и навыками грамотного описания решения задач и представления полученных результатов.

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

1. Постановка задачи предсказания с использованием экспертных стратегий. Понятие (внешнего) регрета. Алгоритм взвешенного большинства. Оценка числа ошибок.
2. Алгоритм оптимального распределения потерь в режиме онлайн. Оценка его регрета.
3. Алгоритм следования за возмущенным лидером. Оценка его регрета. Состоятельность по Ханнану.
4. Задача минимизации регрета с точки зрения теории оптимизации. Алгоритм следования за регуляризованным лидером.
5. Онлайн метод градиентного спуска.
6. Онлайн метод зеркального спуска.
7. Неравенства больших уклонений. Варианты неравенства Хефдинга.
8. Алгоритм экспоненциального взвешивания экспертных решений.
9. Алгоритм экспоненциального взвешивания экспертных решений с переменным параметром обучения.
10. Задача о многоруком бандите. Стохастическая и детерминированная постановки. Алгоритм для детерминированной постановки.
11. Бустинг. Алгоритм Адабуст и его свойства.
12. Агрегирующий алгоритм Вовка. Конечное и бесконечное множество экспертов.
13. Применение агрегирующего алгоритма для различных функций потерь: логарифмической, квадратичной, простой.
14. Универсальный портфель акций Ковера. Алгоритм построения портфеля. Оценка выигрыша.
15. Применение агрегирующего алгоритма для многомерной регрессии в режиме онлайн.
16. Калибруемость предсказаний по Дэвиду. Алгоритмы построения калибруемых предсказаний.
17. Универсальные RKHS. Построение универсальных алгоритмов для онлайн регрессии на основе метода калибруемости.
18. Средние Радемахера и оценка калибровочной ошибки.
19. Агрегирующий алгоритм как результат построения калибруемых предсказаний.

Критерии оценивания

отлично 10 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;

отлично 9 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;

отлично 8 оценка «отлично (8)» выставляется студенту, показавшему всесторонне систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений;

хорошо 7 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;

хорошо 6 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;

хорошо 5 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;

удовлетворительно 4 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;

удовлетворительно 3 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;

неудовлетворительно 2 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

неудовлетворительно 1 оценка «неудовлетворительно (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

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, вычислительной техникой, конспектами лекций.

Дифференцированный зачет может проводиться по итогам текущей успеваемости и сдачи заданий, или путем организации специального опроса, проводимого в устной форме.