

**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**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Методы машинного обучения в прикладных задачах геофизики и биоматематики
<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра вычислительных технологий и моделирования в геофизике и биоматематике
<b>курс:</b>	1
<b>квалификация:</b>	магистр

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

Аудиторных часов: 30 всего, в том числе:

лекции: 0 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час.

Всего часов: 45, всего зач. ед.: 1

Программу составил: С.А. Горейнов, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры вычислительных технологий и моделирования в геофизике и биоматематике 02.04.2022

## Аннотация

В курсе обсуждаются математические основы методов машинного обучения. Основное внимание уделяется алгоритмическим аспектам применения машинного обучения в прикладных задачах геофизической гидродинамики, медицине и биологии.

### 1. Цели и задачи

#### Цель дисциплины

Освоение студентами фундаментальных знаний в области методов машинного обучения, алгоритмическим аспектам применения машинного обучения в прикладных задачах геофизической гидродинамики, медицине и биологии.

#### Задачи дисциплины

- формирование базовых знаний в области методов машинного обучения;
- обучение студентов современным методам машинного обучения;
- формирование подходов к выполнению исследований студентами по математическому моделированию в рамках выпускных работ на степень магистра.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость
	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области математики, естественных наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-4.2 Способен применять знание информационно-коммуникационных технологий для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
	ОПК-4.4 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- место и роль общих вопросов науки в научных исследованиях;
- современные проблемы вычислительной математики;
- новейшие численные методы эффективного решения больших систем, порождаемых задачами математической физики;
- постановку проблем моделирования физических процессов;
- о взаимосвязях и фундаментальном единстве естественных наук.

уметь:

эффективно использовать на практике теоретические компоненты науки: понятия, суждения, умозаключения, законы;  
 представить панораму универсальных методов и законов современного естествознания;  
 работать на современном компьютерном оборудовании;  
 абстрагироваться от несущественных влияний при моделировании реальных физических ситуаций;  
 пользоваться справочной литературой научного и прикладного характера для быстрого поиска необходимых математических и физических данных и понятий.

владеть:

планированием, постановкой и обработкой результатов численного эксперимента;  
 навыками самостоятельной работы в лаборатории на современном компьютерном оборудовании;  
 навыками освоения большого объема информации;  
 культурой постановки и моделирования физических задач.

#### 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	Введение, постановка задач машинного обучения, примеры		10		5
2	Градиентные линейные методы		10		5
3	Линейная регрессия и метод главных компонент		10		5
Итого часов			30		15
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

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

Семестр: 2 (Весенний)

###### 1. Введение, постановка задач машинного обучения, примеры

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.
- Стохастические методы: бэггинг и метод случайных подпространств.

- Нелинейные алгоритмические композиции. Смесь экспертов, область компетентности алгоритма. Выпуклые функции потерь. Методы построения смесей: последовательный и иерархический. Построение смесей экспертов с помощью ЕМ-алгоритма

## 2. Градиентные линейные методы

- Усечённый поиск в ширину, многорядный итерационный алгоритм МГУА.
- Генетический алгоритм, его сходство с МГУА.
- Случайный поиск и Случайный поиск с адаптацией (СПА).

## 3. Линейная регрессия и метод главных компонент

- 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)

Доска, ноутбук и мультимедийное оборудование (проектор или плазменная панель).

## 6. Перечень рекомендуемой литературы

### Основная литература

1. Введение в методы машинного обучения с подкреплением, учебное пособие /А. И. Панов; Министерство науки и высшего образования Российской Федерации ; Московский физико-технический институт (национальный исследовательский университет). Москва, МФТИ, 2019

### Дополнительная литература

1. Машинное обучение [Текст]/Х. Бринк, Дж. Ричардс, М. Феверолф, Real-World Machine Learning, -СПб., Питер, 2017
2. Нетрадиционные методы многомерного статистического анализа [Текст], сб. статей/Б. Эфрон , -М., Финансы и статистика, 1988

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

<http://www.siam.org/>  
<http://sciencedirect.com/>  
<http://ddm.org/>

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

Журналы по научным вычислениям и матричному анализу (напр., SIAM Journal of Scientific Computing <http://www.siam.org>, <http://sciencedirect.com>), доступные через Internet сборники трудов тематических конференций: <http://ddm.org/>, электронные конспекты лекций, учебные пособия и сборники задач, разработанные для данного курса.

Обеспечение самостоятельной работы — электронная библиотека ИВМ РАН.

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

Студент, изучающий дисциплину, 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, алгоритмы.

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;
- подготовку к дифференцированному зачету.

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

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

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра вычислительных технологий и моделирования в геофизике и биоматематике
<b>курс:</b>	<u>1</u>
<b>квалификация:</b>	магистр
Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет	
<b>Разработчик:</b>	С.А. Горейнов, канд. физ.-мат. наук, доцент

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость
	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области математики, естественных наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-4.2 Способен применять знание информационно-коммуникационных технологий для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
	ОПК-4.4 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Методы машинного обучения в прикладных задачах геофизики и биоматематики» обучающийся должен:

### знать:

место и роль общих вопросов науки в научных исследованиях;  
современные проблемы вычислительной математики;  
новейшие численные методы эффективного решения больших систем, порождаемых задачами математической физики;  
постановку проблем моделирования физических процессов;  
о взаимосвязях и фундаментальном единстве естественных наук.

### уметь:

эффективно использовать на практике теоретические компоненты науки: понятия, суждения, умозаключения, законы;  
представить панораму универсальных методов и законов современного естествознания;  
работать на современном компьютерном оборудовании;  
абстрагироваться от несущественных влияний при моделировании реальных физических ситуаций;  
пользоваться справочной литературой научного и прикладного характера для быстрого поиска необходимых математических и физических данных и понятий.

### владеть:

планированием, постановкой и обработкой результатов численного эксперимента;  
навыками самостоятельной работы в лаборатории на современном компьютерном оборудовании;  
навыками освоения большого объема информации;  
культурой постановки и моделирования физических задач.

## 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 прошлой лекции или в конце занятия по пройденной теме.

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

1. Метрические методы классификации: общая формула метрического классификатора, метод ближайших соседей, метод парзеновского окна, метод потенциальных функций и его связь с сетями радиальных базисных функций RBF.
2. Логические методы классификации: понятие логической закономерности, решающие деревья, их построение и усечение, критерии ветвления, бинаризация признаков.
3. Линейные методы классификации: логистическая регрессия, SVM, вероятностная интерпретация функций потерь и регуляризации.
4. Байесовская теория классификации: общая формула оптимального байесовского классификатора, наивный байесовский классификатор, параметрическое и непараметрическое оценивание плотности распределения, EM-алгоритм.

#### **Критерии оценивания**

Оценка отлично 10 баллов - 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.

Оценка отлично 9 баллов - 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 8 баллов - 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.

Оценка хорошо 7 баллов - 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.

Оценка хорошо 6 баллов - 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 5 баллов - 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.

Оценка удовлетворительно 4 балла - 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.

Оценка удовлетворительно 3 балла - 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.

Оценка неудовлетворительно 2 балла - 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.

Оценка неудовлетворительно 1 балл - 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.