

**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**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Оптимизация в машинном обучении
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Компьютерное моделирование физических процессов Физтех-школа Аэрокосмических Технологий кафедра вычислительной физики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Аудиторных часов: 45 всего, в том числе:

лекции: 0 час.

семинары: 45 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составили:

А.В. Гасников, д-р физ.-мат. наук, доцент

Э.А. Горбунов

М.Ю. Данилова

А.В. Рогозин

Программа обсуждена на заседании кафедры вычислительной физики 01.06.2020

## Аннотация

Основная цель курса – познакомить слушателей с основными методами математической оптимизации, а также с современными методами оптимизации, применяемыми в анализе данных и машинном обучении. В частности, мы подробно изучим основные методы первого порядка, способы выбора шага в методах, стохастические методы первого порядка (SGD и его многочисленные вариации), а также их применения к задачам распределённого обучения. В курсе существенное внимание уделяется теоретическим гарантиям сходимости методов для различных задач, вопросам оптимальности методов и их практической эффективности.

### 1. Цели и задачи

#### Цель дисциплины

познакомить студентов с современными методами оптимизации для решения задач, возникающих в машинном обучении, с их теоретическими и практическими свойствами и научить их правильно применять методы оптимизации для решения задач машинного обучения.

#### Задачи дисциплины

1. Познакомить студентов с современными методами оптимизации для решения задач, возникающих в машинном обучении, с их теоретическими и практическими свойствами.
2. Научить студентов правильно применять методы оптимизации для решения задач, в том числе определять наиболее подходящие методы для решения конкретных задач.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности

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

В результате освоения дисциплины обучающиеся должны

знать:

основные виды задач машинного обучения; основные семейства алгоритмов, используемых для решения задач классификации, регрессии, кластеризации; способы отбора признаков, поиска информативных признаков, создания новых признаков; свойства различных метрик качества, функций потерь и способов валидации.

уметь:

самостоятельно решать задачи анализа данных с использованием языка Python; пользоваться математическими пакетами, пакетами визуализации, пакетами глубокого обучения в Python; строить нейронные сети с помощью Keras, Tensorflow, Pytorch.

владеть:

навыками выбора модели, метрики и способа валидации для конкретной задачи; навыками использования Python и внешних библиотек для решения задач анализа данных.

Материально-техническая база: компьютерный класс.

#### 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		2		6
2	Понятие о численных методах оптимизации		4		6
3	Линейные методы и метрические методы		4		6
4	Метод Ньютона		5		6
5	Задачи с регуляризацией, проксимальный оператор		5		6
6	Проксимальный стохастический градиентный спуск (SGD) в общности статьи		5		6
7	Общий взгляд на оптимальные методы невыпуклой оптимизации первого порядка		5		6
8	Задачи оптимизации на множествах простой структуры		5		6
9	Нижние оценки сложности детерминированных методов первого порядка для минимизации гладких/не гладких выпуклых и сильно выпуклых функций, гладких невыпуклых функций		5		6
10	Введение в распределённую оптимизацию		5		6
Итого часов			45		60
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

## 1. Введение

Связь оптимизации и анализа данных, статистики.

## 2. Понятие о численных методах оптимизации

Модель чёрного ящика, сложность задач оптимизации. Классы задач оптимизации. Сильно выпуклые задачи, выпуклые (вырожденные) задачи, невыпуклые задачи. Гладкие, негладкие задачи. Основы матрично-векторного дифференцирования. Условия оптимальности первого и второго порядка, субдифференциал и его свойства.

## 3. Линейные методы и метрические методы

Методы одномерной минимизации (метод дихотомии, метод золотого сечения, метод Фибоначчи). Методы маломерной оптимизации: метод центров тяжести, метод эллипсоидов. Метод градиентного спуска. Способы выбора шага в методах. Наискорейший спуск. Правило Армихо, правило Голдстейна, правило Нестерова. Адаптивный способ выбора шага. Сопряженные направления. Метод сопряженных градиентов для минимизации квадратичных функций. Метод сопряженных градиентов для решения задач выпуклой оптимизации. Метод тяжелого шарика Поляка. Ускоренный градиентный метод (в разных вариантах: метод Нестерова (Nesterov momentum), линейный каплинг, метод подобных треугольников).

## 4. Метод Ньютона

Квазиньютоновские методы (BFGS и LBFGS). Метод Ньютона с кубической регуляризацией.

## 5. Задачи с регуляризацией, проксимальный оператор

Проксимальный градиентный спуск и FISTA.

## 6. Проксимальный стохастический градиентный спуск (SGD) в общности статьи

Методы редукции дисперсии: SVRG, SAGA, Loopless-SVRG как частный случай SGD. Мини-батчинг. Условия сильного (SGC) и слабого (WGC) роста, связь с over-parametrized models. Сходимость ускоренного SGD в предположении SGC.

## 7. Общий взгляд на оптимальные методы невыпуклой оптимизации первого порядка

Невыпуклая оптимизация: общий взгляд на оптимальные методы невыпуклой оптимизации первого порядка (градиентный спуск, SGD и SPIDER). Адаптивные стохастические методы: Momentum SGD, Adagrad, Adadelta, RMSprop, Adam. Новый вариант адаптивного по размеру шага и размер батча SGD.

## 8. Задачи оптимизации на множествах простой структуры

Дивергенция Брэгмана. Метод проекции градиента, метод зеркального спуска. Метод условного градиента (Франк-Вульфа). Пример задачи минимизации квадратичной формы с разреженной положительно определенной матрицей на единичном симплексе.

## 9. Нижние оценки сложности детерминированных методов первого порядка для минимизации гладких/не гладких выпуклых и сильно выпуклых функций, гладких невыпуклых функций

NP-сложность задач невыпуклой негладкой оптимизации, связь с NP-сложностью задачи о камнях. Нижние оценки глобальной оптимизации, концепция сопротивляющегося оракула. Нижние оценки сложности в стохастической оптимизации первого порядка, нижние оценки сложности методов первого порядка для минимизации сумм. Оптимальные методы для минимизации сумм: Katyusha, MiG, SSNM (гладкие выпуклый и сильно выпуклый случаи). Оптимальные методы для общей задачи стохастической оптимизации: стохастический вариант метода Нестерова с мини-батчингом (гладкий выпуклый и сильно выпуклый случаи), стохастический зеркальный спуск (негладкий выпуклый и сильно выпуклый случаи).

## 10. Введение в распределённую оптимизацию

Централизованная и децентрализованная оптимизация на примере SGD. Способы уменьшения затрат на коммуникации: спарсификация, квантизация и федеративное обучение.

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

Компьютер и мультимедийное оборудование (проектор).

## 6. Перечень рекомендуемой литературы

### Основная литература

1. Стохастический анализ в задачах [Текст]: Ч.1 / под ред. А. В. Гасникова; М-во образования и науки РФ, Моск. физ.-техн. ин-т (гос. ун-т) - ММФТИ, 2016
  2. Введение в оптимизацию [Текст] / Б. Т. Поляк - М. ЛЕНАНД, 2014
- Фонд литературы базовой кафедры

1. Гасников А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска. – М.: МФТИ, 2018. – 241 с. 2-е изд.  
<https://arxiv.org/ftp/arxiv/papers/1711/1711.00394.pdf>
2. Презентации к некоторым частям курса доступны по ссылке (наиболее важными являются презентации 1-4) <https://www.mccme.ru/dubna/2017/courses/gasnikov.html>
3. Boyd S., Vandenberghe L. Convex optimization. – Cambridge University Press, 2004.
4. Bubeck S. Convex optimization: algorithms and complexity // Foundations and Trends in Machine Learning. – 2015. – V. 8, N 3–4. – P. 231–357.
5. Nemirovski A. Advanced Nonlinear Programming // Lectures, ISyE 7683 Spring 2019. – URL: [https://www2.isye.gatech.edu/~nemirovs/Trans\\_ModConvOpt.pdf](https://www2.isye.gatech.edu/~nemirovs/Trans_ModConvOpt.pdf)
6. Nesterov Yu. Lectures on convex optimization. – Springer, 2018.

### Дополнительная литература

1. Численные методы решения экстремальных задач [Текст] : учеб. пособие / Ф.П.Васильев .— 2-е изд., перераб. и доп. — М. : Наука, 1988 .— 549 с.
  2. Курс методов оптимизации [Текст]/А. Г. Сухарев, А. В. Тимохов, В. В. Федоров, -М., Наука, 1986
  3. Численные методы оптимизации [Текст] : [учеб.пособие для вузов] / А.Ф.Измайлов, М.В.Солодов .— М. : Физматлит, 2003, 2005 .— 304 с.
- Фонд литературы базовой кафедры

1. Nocedal J., Wright S. Numerical optimization. – Springer, 2006.
2. Евтушенко Ю.Г. Оптимизация и быстрое автоматическое дифференцирование. – М.: ВЦ РАН, 2013. – 144 с. – URL: <http://www.ccas.ru/personal/evtush/p/198.pdf>
3. Гасников А.В. Оптимизация и анализ данных // Курс лекций в Адыгейском государственном университете. – Кавказский математический центр, 26–30 ноября 2018 г. – URL: <https://www.youtube.com/playlist?list=PLVdKoATzj9WQM9kWYzdED3aBVW3FrnU64>
4. <https://github.com/amkatrutsa/MIPT-Opt>
5. [http://nbviewer.jupyter.org/github/merkulovdaniil/mipt\\_optimization/tree/master/](http://nbviewer.jupyter.org/github/merkulovdaniil/mipt_optimization/tree/master/)  
<https://www.youtube.com/playlist?list=PLQSHEO58cjmMuDtDH0cu3ASh7QIVUiaWG> Видео лекций по аналогичному курсу 2018 года

**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 занятиях используются мультимедийные технологии, включая демонстрацию презентаций. Дистрибутив Anaconda.

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

Методические рекомендации позволяют студенту оптимальным образом организовать процесс обучения. В рабочей программе приведено примерное распределение часов аудиторной и внеаудиторной нагрузки по различным темам данной дисциплины.

Для успешного освоения данной дисциплины студенту необходимо посещать лекции и семинарские занятия,

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Компьютерное моделирование физических процессов Физтех-школа Аэрокосмических Технологий кафедра вычислительной физики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

**Разработчики:**

А.В. Гасников, д-р физ.-мат. наук, доцент

Э.А. Горбунов

М.Ю. Данилова

А.В. Рогозин

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Оптимизация в машинном обучении» обучающийся должен:

### знать:

основные виды задач машинного обучения; основные семейства алгоритмов, используемых для решения задач классификации, регрессии, кластеризации; способы отбора признаков, поиска информативных признаков, создания новых признаков; свойства различных метрик качества, функций потерь и способов валидации.

### уметь:

самостоятельно решать задачи анализа данных с использованием языка Python; пользоваться математическими пакетами, пакетами визуализации, пакетами глубокого обучения в Python; строить нейронные сети с помощью Keras, Tensorflow, Pytorch.

### владеть:

навыками выбора модели, метрики и способа валидации для конкретной задачи; навыками использования Python и внешних библиотек для решения задач анализа данных.

Материально-техническая база: компьютерный класс.

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

Типовые контрольные задания

1. Связь оптимизации и анализа данных, статистики
2. Понятие о численных методах оптимизации. Модель чёрного ящика, сложность задач оптимизации. Сильно выпуклые задачи, выпуклые (вырожденные) задачи, невыпуклые задачи. Гладкие, негладкие задачи. Основы матрично-векторного дифференцирования. Условия оптимальности первого и второго порядка, субдифференциал и его свойства.



3. Унимодальные функции одной переменной. Методы одномерной минимизации (метод дихотомии, метод золотого сечения, метод Фибоначчи). Методы маломерной оптимизации: метод центров тяжести, метод эллипсоидов. Метод градиентного спуска. Способы выбора шага в методах. Наискорейший спуск. Правило Армихо, правило Голдстейна, правило Нестерова. Адаптивный способ выбора шага. Сопряженные направления. Метод сопряженных градиентов для минимизации квадратичных функций. Метод сопряженных градиентов для решения задач выпуклой оптимизации. Метод тяжелого шарика Поляка. Ускоренный градиентный метод (в разных вариантах: метод Нестерова (Nesterov momentum), линейный каплинг, метод подобных треугольников).
4. Задачи оптимизации на множествах простой структуры. Дивергенция Брэгмана. Метод проекции градиента, метод зеркального спуска. Метод условного градиента (Франк-Вульфа). Пример задачи минимизации квадратичной формы с разреженной положительно определенной матрицей на единичном симплексе. Задачи с регуляризацией, проксимальный оператор. Проксимальный градиентный спуск и FISTA.
5. Проксимальный стохастический градиентный спуск (SGD) в общности статьи. Методы редукции дисперсии: SVRG, SAGA, Loopless-SVRG как частный случай SGD. Мини-батчинг. Условия сильного (SGC) и слабого (WGC) роста, связь с over-parametrized models. Сходимость ускоренного SGD в предположении SGC.
6. Нижние оценки сложности детерминированных методов первого порядка для минимизации гладких/не гладких выпуклых и сильно выпуклых функций, гладких невыпуклых функций. NP-сложность задач невыпуклой негладкой оптимизации, связь с NP-сложностью задачи о камнях (с обсуждением деталей об экспоненциальности весов). Нижние оценки глобальной оптимизации, концепция сопротивляющегося оракула. Нижние оценки сложности в стохастической оптимизации первого порядка, нижние оценки сложности методов первого порядка для минимизации сумм. Оптимальные методы для минимизации сумм: Katyusha, MiG, SSNM (гладкие выпуклый и сильно выпуклый случаи), (гладкий невыпуклый случай). Оптимальные методы для общей задачи стохастической оптимизации: стохастический вариант метода Нестерова с мини-батчингом (гладкий выпуклый и сильно выпуклый случаи), стохастический зеркальный спуск (негладкий выпуклый и сильно выпуклый случаи). Оптимальность SGD в гладком невыпуклом случае.
7. Невыпуклая оптимизация: общий взгляд на оптимальные методы невыпуклой оптимизации первого порядка. Адаптивные стохастические методы: Momentum SGD, Adagrad, Adadelta, RMSprop, Adam (изложение на уровне обзора Себастьяна Рюдера; недостатки Adadelta, Adam и RMSprop. Новый вариант адаптивного по размеру шага и размер батча SGD.
8. Введение в распределённую оптимизацию: централизованная и децентрализованная оптимизация на примере SGD. Способы уменьшения затрат на коммуникации: спарсификация, квантизация и федеративное обучение.
9. Метод Ньютона. Квазиньютоновские методы (BFGS и LBFGS). Метод Ньютона с кубической регуляризацией

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

##### **Примеры типовых вопросов**

1. Понятие о численных методах оптимизации
2. Линейные методы и метрические методы
3. Метод Ньютона
4. Задачи с регуляризацией, проксимальный оператор
5. Проксимальный стохастический градиентный спуск (SGD) в общности статьи

##### **Примеры билетов**

1. Общий взгляд на оптимальные методы невыпуклой оптимизации первого порядка
2. Задачи оптимизации на множествах простой структуры
3. Нижние оценки сложности детерминированных методов первого порядка для минимизации гладких/не гладких выпуклых и сильно выпуклых функций, гладких невыпуклых функций
4. Введение в распределённую оптимизацию

##### **Критерии оценивания**

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) - выставляется студенту, показавшему свободное оперирование знаниями учебной программы дисциплины, выполнение заданий творческого характера.

отлично (8) - выставляется студенту, показавшему владение программным учебным материалом с наличием несущественных ошибок в действиях, самостоятельно исправляемых учащимся.

хорошо (7) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускается в ответе или в решении задач некоторые неточности.

хорошо (6) - выставляется студенту если он осознает воспроизведение программного учебного материала, в том числе и различной степени сложности, с несущественными ошибками, затруднения в применении отдельных навыков.

хорошо (5) - выставляется студенту если теоретическое содержание освоено не полностью, некоторые практические навыки сформированы недостаточно, в некоторых случаях были допущены ошибки.

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) - выставляется студенту в случае большого количества недочетов и неправильных ответов, а также пассивной работе в ходе занятий, многие учебные задания не выполнены.

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) - выставляется студенту, который не освоил теоретическое и практическое содержание курса, все выполненные учебные задания содержат грубые ошибки.

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

Во время проведения оценивания обучающиеся могут пользоваться программой дисциплины, а также конспектами лекций.

Оценка выставляется путем организации специального опроса, проводимого в устной форме (время на подготовку не более 60 минут, время на ответ не более 30 минут).