

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

**УТВЕРЖДЕНО**  
**Директор**

**А.В. Малеев**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Основы машинного обучения
<b>по направлению:</b>	Программная инженерия
<b>профиль подготовки:</b>	Разработка программно-информационных систем высшая школа программной инженерии высшая школа программной инженерии МФТИ - Яндекс
<b>курс:</b>	2
<b>квалификация:</b>	бакалавр

Семестры, формы промежуточной аттестации:

3 (осенний) - Дифференцированный зачет

4 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 84 час.

Всего часов: 144, всего зач. ед.: 4

Программу составил: А.В. Малеев, директор

Программа обсуждена на заседании высшей школы программной инженерии МФТИ - Яндекс 05.03.2022

## Аннотация

Курс предназначен для магистров математики, интересующихся современными методами дискретной математики и вероятностными приложениями в информатике. Курс включает в себя все основные определения и положения теории вероятностей: аксиоматику Колмогорова, распределения, случайные величины и векторы, математическое ожидание, вероятностные неравенства, законы больших чисел и центральную предельную теорему.

### 1. Цели и задачи

#### Цель дисциплины

овладение основными современными методами теории вероятностей.

#### Задачи дисциплины

- студенты овладевают базовыми знаниями (понятиями, концепциями, методами и моделями) теории вероятностей;
- приобретение теоретических знаний и практических навыков по теории вероятностей;
- консультирование и помощь студентам в проведении собственных теоретических исследований по теории вероятностей.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.3 Способен определять границы применимости полученных результатов
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-2.1 Способен использовать информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	ОПК-2.2 Владеет навыками создания программного обеспечения для ЭВМ и систем различной архитектуры

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

В результате освоения дисциплины обучающиеся должны

знать:

- ☐ фундаментальные понятия, законы теории вероятностей;
- ☐ современные проблемы соответствующих разделов теории вероятностей;
- ☐ понятия, аксиомы, методы доказательства и доказательства основных теорем в разделах, входящих в основную часть цикла;
- ☐ основные свойства соответствующих математических объектов;
- ☐ аналитические и численные подходы и методы решения типовых прикладных задач теории вероятностей.

уметь:

- ☐ понять задачу;
- ☐ использовать свои знания для решения фундаментальных и прикладных задач;
- ☐ оценить правильность постановки задачи;
- ☐ строго доказывать или опровергать заявление;
- ☐ самостоятельно находить алгоритмы решения проблем, в том числе нестандартных, и проводить их анализ;
- ☐ самостоятельно видеть последствия результатов;
- ☐ точно представлять математические знания по теории вероятностей в устной и письменной форме.

владеть:

- ☐ навыки усвоения большого количества информации и решения задач (в том числе сложных);
- ☐ навыки самостоятельной работы и освоения новых дисциплин;
- ☐ культура постановки, анализа и решения математических и прикладных задач, требующих использования математических подходов и методов их решения;
- ☐ предметный язык теории вероятностей и навыки грамотного описания решения задач и представления 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	Введение в машинное обучение. Метрические алгоритмы, оценка качества моделей	4	4		14
2	Линейные модели	5	5		14
3	Деревья и ансамбли моделей	6	6		14
4	Работа с признаками. Ограничения машинного обучения	4	4		14
5	Введение в глубокое обучение	5	5		14
6	Обучение без учителя	6	6		14
Итого часов		30	30		84
Подготовка к экзамену		0 час.			
Общая трудоёмкость		144 час., 4 зач.ед.			

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

###### Семестр: 3 (Осенний)

###### 1. Дискретные вероятностные пространства.

Дискретные вероятностные пространства. Классическое определение вероятности. Примеры.

###### 2. Независимость от произвольного набора случайных величин.

Независимость от произвольного набора случайных величин. Критерий независимости, теорема о независимости борелевских функций от непересекающихся множеств независимых случайных величин.

###### 3. Случайные элементы, случайные величины и векторы.

Случайные элементы, случайные величины и векторы. Достаточное условие измеримости отображения, следствие для случайных величин и векторов. Действия со случайными величинами.

###### Семестр: 4 (Весенний)

###### 4. Теорема Каратеодори о продолжении вероятностной меры (доказательство единственности).

Теорема Каратеодори о продолжении вероятностной меры (доказательство единственности).  
Теорема Лебега о функции распределения

## 5. Условные вероятности.

Условные вероятности. Формула полной вероятности. Формула Байеса. Примеры

## 6. Обучение без учителя

Кластеризация. Метрический подход, алгоритм k-means. Иерархическая кластеризация. Алгоритм DBSCAN.

Методы снижения размерности. Многомерное шкалирование. Isomap. Locally Linear Embedding. SNE, t-SNE.

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

Стандартная учебная аудитория

## **6.Перечень рекомендуемой литературы**

### Основная литература

1. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль . — Москва, ДМК Пресс, 2018.— URL: <https://e.lanbook.com/book/107901> (дата обращения: 29.01.2021). - Полный текст (Режим доступа : из сети МФТИ / Удаленный доступ)
2. Математические основы машинного обучения и прогнозирования, Электронная версия печатной публикации / В. В. Вьюгин. — Москва, МЦНМО, 2014

### Дополнительная литература

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

<http://dm.fizteh.ru/>

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

Мультимедийные технологии можно использовать на лекциях и практических занятиях, в том числе на презентациях.

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

1. Рекомендуется успешно сдать контрольные работы, так как это упрощает итоговую аттестацию по предмету.
2. Для подготовки к итоговой аттестации по предмету лучше всего использовать лекционные материалы.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Программная инженерия
<b>профиль подготовки:</b>	Разработка программно-информационных систем высшая школа программной инженерии МФТИ - Яндекс высшая школа программной инженерии
<b>курс:</b>	2
<b>квалификация:</b>	бакалавр

Семестры, формы промежуточной аттестации:

- 3 (осенний) - Дифференцированный зачет
- 4 (весенний) - Дифференцированный зачет

**Разработчик:** А.В. Малеев, директор

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.3 Способен определять границы применимости полученных результатов
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-2.1 Способен использовать информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	ОПК-2.2 Владеет навыками создания программного обеспечения для ЭВМ и систем различной архитектуры

## 2. Показатели оценивания компетенций

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

### знать:

- ☐ фундаментальные понятия, законы теории вероятностей;
- ☐ современные проблемы соответствующих разделов теории вероятностей;
- ☐ понятия, аксиомы, методы доказательства и доказательства основных теорем в разделах, входящих в основную часть цикла;
- ☐ основные свойства соответствующих математических объектов;
- ☐ аналитические и численные подходы и методы решения типовых прикладных задач теории вероятностей.

### уметь:

- ☐ понять задачу;
- ☐ использовать свои знания для решения фундаментальных и прикладных задач;
- ☐ оценить правильность постановки задачи;
- ☐ строго доказывать или опровергать заявление;
- ☐ самостоятельно находить алгоритмы решения проблем, в том числе нестандартных, и проводить их анализ;
- ☐ самостоятельно видеть последствия результатов;
- ☐ точно представлять математические знания по теории вероятностей в устной и письменной форме.

### владеть:

- ☐ навыки усвоения большого количества информации и решения задач (в том числе сложных);
- ☐ навыки самостоятельной работы и освоения новых дисциплин;
- ☐ культура постановки, анализа и решения математических и прикладных задач, требующих использования математических подходов и методов их решения;
- ☐ предметный язык теории вероятностей и навыки грамотного описания решения задач и представления результатов.

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

Примеры домашних заданий

- Точка  $x$  выбирается случайным образом из прямоугольника со сторонами длиной 1 и 2. Найдите вероятность того, что ближайшая к  $x$  сторона имеет длину 1.
- $2I(x + y \leq 1, x > 0, y > 0)$  - плотность случайного вектора  $(\xi, \eta)$ . Пусть  $p_\xi, p_\eta, p_{\xi + \eta}$  - плотности (непрерывные на  $(0,1)$ )  $\xi, \eta, \xi + \eta$  соответственно. Найдите  $p_\xi(1/2) p_\eta(1/2) p_{\xi + \eta}(1/2)$ .
- Пусть  $\xi \sim \text{Bern}(p)$  и  $F$  - его функция распределения. Какое распределение имеет  $F(\xi)$ ?
- Пусть  $S_n, n \in \mathbb{Z}^+$  - симметричное случайное блуждание на  $\mathbb{Z}$ . Пусть  $X$  - число  $k \in \{1, \dots, 10\}$  таких, что  $S_k = 0$ . Найдите  $\text{Var} X$ .
- Пусть  $(\xi_n - \xi) 2^n \rightarrow 0$ . Верно ли, что а)  $\xi_n \rightarrow \xi$ ? б)  $\xi_{2^n} \rightarrow \xi$ ? в)  $\xi_{3^n} \rightarrow \xi$ ?

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

Перечень контрольных вопросов к экзамену:

1. Вероятностное пространство. Аксиомы Колмогорова. Теорема о непрерывности в «нуле» вероятностной меры.
2. Дискретные вероятностные пространства. Классическое определение вероятности. Примеры.
3. Геометрические вероятности. Примеры.
4. Условные вероятности. Формула полной вероятности. Примеры формул Байеса.
5. Теорема о монотонных классах.
6. Независимость событий и систем событий. Пример Бернштейна. Лемма о достаточном условии независимости сигма-алгебр.
7. Теорема Каратеодори о продолжении вероятностной меры (доказательство единственности).
8. Случайные величины в дискретных вероятностных пространствах. Независимость от случайных величин. Математическое ожидание случайной величины, ее основные свойства. Дисперсия, ковариантность и их свойства.
9. Случайные элементы, случайные величины и векторы. Достаточное условие измеримости отображения, следствие для случайных величин и векторов. Действия со случайными величинами.
10. Независимость от произвольного набора случайных величин. Критерий независимости, теорема о независимости борелевских функций от непересекающихся множеств независимых случайных величин.
11. Математическое ожидание случайной величины (интеграл Лебега в вероятностной мере): определение простых, неотрицательных и произвольных случайных величин. Проверка определений.
12. Основные свойства математического ожидания. Теорема о математическом ожидании для произведения независимых случайных величин.
13. Дисперсия, ковариантность и их свойства. Неравенство Коши-Буняковского. Дисперсия суммы независимых случайных величин. Ковариационная матрица случайного вектора, ее неотрицательная определенность.
14. Неравенство Маркова, неравенство Чебышева. Закон больших чисел в форме Чебышева. Неравенство Дженсена.
15. Типы сходимости случайных величин, их взаимосвязь. Критерий Коши сходимости с вероятностью 1.
16. Неравенство Колмогорова. Теорема сходимости для почти наверняка серии случайных величин.
17. Усиленный закон больших чисел для независимых случайных величин с ограниченными дисперсиями.
18. Предельный переход под знаком интеграла Лебега. Теорема о монотонной сходимости, лемма Фату, теорема Лебега о мажорированной сходимости.
19. Лемма Бореля Кантелли. Усиленный закон больших чисел для независимых одинаково распределенных случайных величин с ограниченным математическим ожиданием.
20. Формула пересчета математических ожиданий. Теорема о замене переменной в интеграле Лебега.
21. Прямое произведение вероятностных пространств. Теорема Фубини (b / d). Совместное распределение конечного набора случайных величин. Свертка распределений.
22. Слабая сходимость и сходимость в основном вероятностных мер. Теорема Александрова (б / г). Теорема об эквивалентности сходимости по распределению случайных величин и сходимости функций распределения во всех точках непрерывности предельной функции.
23. План тестов Бернулли и полиномиальный план. Предельные теоремы для схемы Бернулли: теорема Пуассона и теорема Муавра-Лапласа (b / d).
24. Характеристические функции вероятностных мер. функции распределения, случайные величины и векторы. Примеры. Основные свойства характеристических функций случайных величин.
25. Теорема единственности для характеристических функций вероятности. Независимость компонент случайного вектора в терминах характеристических функций.

26. Теорема непрерывности для характеристических функций (b / d). Центральная предельная теорема для независимых одинаково распределенных случайных величин.

27. Теорема Берри Эссеена о скорости сходимости в центральной предельной теореме (b / d). Оценки константы в теореме Берри Эссеена.

Билет 1

1. Дизайн тестов Бернулли и полиномиальный дизайн. Предельные теоремы для схемы Бернулли: теорема Пуассона и теорема Муавра-Лапласа (b / d)

2. Неравенство Маркова, неравенство Чебышева. Закон больших чисел в форме Чебышева. Неравенство Дженсена

#### Критерии оценивания

Оценка «отлично (10)» выставляется студенту, проявившему всестороннее, систематическое и глубокое знание материала образовательной программы, самостоятельно выполнившему все задания, предусмотренные программой, глубоко изучившему основную и дополнительную литературу, рекомендованную программой. , активно работает в классе и понимает основные научные концепции по изучаемой дисциплине, проявил творческий подход и научный подход в понимании и представлении материала образовательной программы, ответ на который характеризуется использованием богатых и адекватных терминов, а также последовательным и логичным изложением материала;

Оценка «отлично (9)» выставляется студенту, который продемонстрировал всестороннее систематическое знание материала образовательной программы, самостоятельно выполнил все задачи, предусмотренные программой, глубоко усвоил основную литературу и знаком с рекомендуемой дополнительной литературой. по программе, активно проработал на занятиях, показал системность знаний по дисциплине, достаточную для дальнейшего изучения, а также умение самостоятельно расширять ее, ответ которой отличается точностью используемых терминов, а изложение материала в нем последовательное и логичное;

Оценка «отлично (8)» выставляется студенту, который проявил полное знание материала образовательной программы, не допускает существенных неточностей в своем ответе, самостоятельно выполнил все задания, предусмотренные программой, изучил основную литературу, рекомендованную учебной программой. программа, активно проработанная на занятиях, показала системность его знаний по дисциплине, достаточных для дальнейшего изучения, а также способность самостоятельно их расширять;

Оценка «хорошо (7)» выставляется студенту, который проявил достаточно полное знание материала образовательной программы, не допускает существенных неточностей в ответе, самостоятельно выполнил все задания, предусмотренные программой, изучил основную рекомендованную литературу по программе, активно работал на занятиях, показал системность своих знаний по дисциплине, достаточных для дальнейшего изучения, а также умение самостоятельно их усиливать;

Оценка «хорошо (6)» выставляется студенту, который проявил достаточно полное знание материала образовательной программы, не допускает существенных неточностей в своем ответе, самостоятельно выполнил основные задачи, предусмотренные программой, изучил основную литературу. рекомендован программой, показал систематичность своих знаний по дисциплине, достаточную для дальнейшего изучения;

Оценка «хорошо (5)» дается студенту, продемонстрировавшему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, который, не проявляя достаточной активности на уроках, тем не менее самостоятельно выполнял овладел основными задачами, предусмотренными программой, освоил основную литературу, рекомендованную программой, допустил ошибки в их выполнении и ответе во время тестирования, но имеет необходимые знания для исправления этих ошибок самостоятельно;

Оценка «удовлетворительно (4)» дается студенту, обнаружившему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, который, не проявляя достаточной активности на уроках, тем не менее самостоятельно выполнял основные задачи, предусмотренные программой, изучил основную литературу, но допустил ошибки в их выполнении и в своем ответе во время теста, но имеет необходимые знания для исправления этих ошибок под руководством преподавателя;



Оценка «удовлетворительно (3)» выставляется обучающемуся, проявившему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, не проявившего активности на занятиях, самостоятельно выполнившего основные задачи, предусмотренные программой, но допускающая ошибки в их выполнении и в ответе при тестировании, но обладающая необходимыми знаниями для устранения под руководством преподавателя наиболее существенных ошибок;

Оценка «неудовлетворительно (2)» дается студенту, который показал пробелы в знаниях или недостаток знаний по значительной части материала основной образовательной программы, не выполнил самостоятельно основные задачи, требуемые программой, допустил принципиальные ошибки в выполнении предусмотренных программой задач, который не может продолжить учебу или начать профессиональную деятельность без дополнительной подготовки по рассматриваемой дисциплине;

Оценка «неудовлетворительно (1)» ставится студенту при отсутствии ответа (отказ от ответа), либо когда представленный ответ совсем не соответствует сути вопросов, содержащихся в задании.

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

Во время экзамена студенту разрешается использовать программу дисциплины.