

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

**УТВЕРЖДЕНО**  
**Проректор по учебной работе**

**А.А. Воронов**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Основные методы кластеризации и распознавания
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Радиотехника и компьютерные технологии Физтех-школа Радиотехники и Компьютерных Технологий кафедра проблем передачи информации и анализа данных
<b>курс:</b>	4
<b>квалификация:</b>	бакалавр

Семестр, формы промежуточной аттестации: 7 (осенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: С.М. Карпенко, преподаватель

Программа обсуждена на заседании кафедры проблем передачи информации и анализа данных 15.05.2023

## Аннотация

В курсе рассматриваются ключевые понятия и современные математические методы машинного обучения, а также их приложения в различных областях.

### 1. Цели и задачи

#### Цель дисциплины

Изучение современных алгоритмов обучения машин и распознавания образов.

#### Задачи дисциплины

- подготовка к участию в научных семинарах, научно-технических конференциях и симпозиумах, составлению научных обзоров, рефератов и библиографии по тематике исследований;
- подготовка к оказанию консалтинговых услуг по данной тематике;
- подготовка к участию в международных проектах по тематике дисциплины;
- подготовка к участию в разработке корпоративной политики и мероприятиях в области повышения социальной ответственности бизнеса перед обществом, включая разработку и реализацию решений, направленных на поддержку социально-значимых проектов;
- совершенствование и расширение общенаучной базы. Повышение уровня общекультурного и нравственного совершенствования своей личности.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
ПК-3 Способен выбирать и применять подходящее оборудование, инструменты и методы исследований для решения задач в избранной предметной области	ПК-3.2 Знает области и критерии применимости используемых теоретических подходов и умение оценивать точность приближенных аналитических методов вычислений
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.2 Знает источники происхождения и умеет производить оценку погрешности измерений и достоверности экспериментальных результатов

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

В результате освоения дисциплины обучающиеся должны

знать:

- методологию и терминологию дисциплины;
- механизмы формирования, представления и искажения изображений;
- принципы построения алгоритмов обработки изображений;
- стандартные методы синтеза, восстановления, анализа, классификации и распознавания изображений.

уметь:

- использовать новые знания и применять их в профессиональной деятельности;
- использовать современные теории, методы, системы и средства прикладной математики и информационных технологий для решения научно-исследовательских и прикладных задач.

владеть:

- основами методологии научного познания и системного подхода при изучении различных уровней организации материи, информации, пространства и времени.

### 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. Основные понятия и базовые теоремы.	4	3		8
2	Вероятностный вывод.	8	3		7
3	Обучение без учителя и анализ данных.	8	10		8
4	Обучение с учителем.	10	14		7
Итого часов		30	30		30
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

##### Семестр: 7 (Осенний)

##### 1. Введение. Основные понятия и базовые теоремы.

Обучаемые (параметризованные) алгоритмы. Приложения. Данные, признаки. Обзор протоколов обучения: с учителем, без учителя, с подкреплением. Примеры.

Роль методов оптимизации. Нейронные сети, коннекционизм. Переобучение и регуляризация.

##### 2. Вероятностный вывод.

Скрытые Марковские модели. Марковские случайные поля. Общий взгляд на модели со скрытыми параметрами.

##### 3. Обучение без учителя и анализ данных.

Кластеризация. Иерархическая кластеризация. Алгоритм К-средних. Модель смеси гауссиан. Дискриминант Фишера. Алгоритм ожидания-максимизации (ЕМ).

Сокращение размерности. Анализ главных компонент (PCA). Нейронные сети, осуществляющие нелинейный анализ главных компонент.

Пропущенные данные. Вероятностная трактовка PCA.

Факторный анализ, анализ независимых компонент.

##### 4. Обучение с учителем.

Логистическая регрессия, персептрон. Обратное распространение ошибки. Многослойный персептрон. Разделяемые веса. Обучаемые метрики (сиамские сети). Конволютивные сети. Машины опорных векторов (SVM). Квадратичная оптимизация. Нестандартные скалярные произведения.

Простейший алгоритм обучения, использующий теорему Байеса.

Практические вопросы: сбор базы данных, выбор признаков, диагностика качества работы алгоритма. Типы ошибок, характеристическая кривая (ROC-curve).

Деревья принятия решений, Алгоритм C4.5. Бустинг. Алгоритм Виолы-Джонса.

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

Компьютерный класс, доска, ноутбук и мультимедийное оборудование (проектор или плазменная панель).

## **6.Перечень рекомендуемой литературы**

### **Основная литература**

1. Теория распознавания образов : Статистические методы [Текст] : учеб. пособие для вузов / А. А. Натан ; М-во высш. и средн. спец. образования РСФСР, Моск. физ.-техн. ин-т .— М. : МФТИ, 1988 .— 84 с. : ил. - 400 экз.

фонд литературы кафедры

2. Мерков А.Б. Введение в методы статистического обучения // Материалы к учебнику: <http://www.recognition.mccme.ru/pub/RecognitionLab.html/sltb.pdf> .

3. Bishop C. Pattern Recognition and Machine Learning. Springer, 2006. – 101 p. [имеется в библиотечном фонде кафедры]

4. Ветров Д.П., Кропотов Д.А. Алгоритмы выбора моделей и построения коллективных решений в задачах классификации, основанные на принципе устойчивости. М.: КомКнига, 2006. – 112 с.

5. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. – 511 с.

### **Дополнительная литература**

1. Цифровая обработка изображений, Электронная версия печатной публикации / Р. Гонсалес, Р. Вудс. — Москва, Техносфера, 2012

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

<http://www.recognition.mccme.ru/pub/RecognitionLab.html/sltb.pdf>

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

Программное обеспечение и информационные технологии не требуются.

## **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, 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;
- выполнение лабораторных работ, для осознания связей между теорией и практическими навыками.
- 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>	Прикладные математика и физика
<b>профиль подготовки:</b>	Радиотехника и компьютерные технологии Физтех-школа Радиотехники и Компьютерных Технологий кафедра проблем передачи информации и анализа данных
<b>курс:</b>	4
<b>квалификация:</b>	бакалавр
Семестр, формы промежуточной аттестации: 7 (осенний) - Дифференцированный зачет	
<b>Разработчик:</b>	С.М. Карпенко, преподаватель

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
ПК-3 Способен выбирать и применять подходящее оборудование, инструменты и методы исследований для решения задач в избранной предметной области	ПК-3.2 Знает области и критерии применимости используемых теоретических подходов и умение оценивать точность приближенных аналитических методов вычислений
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.2 Знает источники происхождения и умеет производить оценку погрешности измерений и достоверности экспериментальных результатов

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Основные методы кластеризации и распознавания» обучающийся должен:

### знать:

- методологию и терминологию дисциплины;
- механизмы формирования, представления и искажения изображений;
- принципы построения алгоритмов обработки изображений;
- стандартные методы синтеза, восстановления, анализа, классификации и распознавания изображений.

### уметь:

- использовать новые знания и применять их в профессиональной деятельности;
- использовать современные теории, методы, системы и средства прикладной математики и информационных технологий для решения научно-исследовательских и прикладных задач.

### владеть:

- основами методологии научного познания и системного подхода при изучении различных уровней организации материи, информации, пространства и времени.

## 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. Обучаемые (параметризованные) алгоритмы. Приложения. Данные, признаки.
2. Обзор протоколов обучения: с учителем, без учителя, с подкреплением. Примеры.
3. Роль методов оптимизации. Нейронные сети, коннекционизм. Переобучение и регуляризация.
4. Логистическая регрессия, персептрон.
5. Обратное распространение ошибки. Многослойный персептрон. Разделяемые веса. Обучаемые метрики (сиамские сети). Конволютивные сети.
6. Машины опорных векторов (SVM). Квадратичная оптимизация. Нестандартные скалярные произведения.
7. Простейший алгоритм обучения, использующий теорему Байеса. Практические вопросы: сбор базы данных, выбор признаков, диагностика качества работы алгоритма. Типы ошибок, характеристическая кривая (ROC-curve).
8. Деревья принятия решений, Алгоритм C4.5.
9. Бустинг. Алгоритм Виолы-Джонса.

10. Кластеризация. Иерархическая кластеризация. Алгоритм К-средних. Модель смеси гауссиан. Дискриминант Фишера.
11. Алгоритм ожидания-максимизации (ЕМ).
12. Сокращение размерности. Анализ главных компонент (РСА). Нейронные сети, осуществляющие нелинейный анализ главных компонент.
13. Пропущенные данные. Вероятностная трактовка РСА.
14. Скрытые Марковские модели.

Примерный перечень билетов:

Билет №1.

1. Обучаемые (параметризованные) алгоритмы. Приложения. Данные, признаки.
2. Роль методов оптимизации. Нейронные сети, коннекционизм. Переобучение и регуляризация.

Билет №2.

1. Обзор протоколов обучения: с учителем, без учителя, с подкреплением. Примеры.
2. Логистическая регрессия, персептрон.

### 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 проводится путем организации специального опроса, проводимого в устной форме.