
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол №2

**Рабочая программа дисциплины (модуля)
«Статистическая теория обучения»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Прикладной искусственный интеллект

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2026

**Москва
2026**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения	5
3. Тематический план	7
4. Содержание дисциплины (модуля)	8
5. Учебно-методическое обеспечение	9
6. Материально-техническое обеспечение	9
7. Методические и оценочные материалы	11

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Статистическая теория обучения» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Дисциплина (модуль) «Статистическая теория обучения» позволяет понимать, почему и при каких условиях алгоритмы обучаются, как оценивать их обобщающую способность и избегать переобучения. Эта дисциплина критически важна для подготовки специалистов, способных не только применять, но и осмысливать, модифицировать и разрабатывать эффективные алгоритмы обучения.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) доступна для изучения на 3 или 4 курсе в 6, 7, 8 семестрах на выбор после успешного освоения дисциплин (модулей): «Математический анализ. Часть 1», «Математический анализ. Часть 2» и «Линейная алгебра и геометрия. Часть 1», «Линейная алгебра и геометрия. Часть 2» или «Математический анализ. Продвинутый уровень. Часть 1», «Математический анализ. Продвинутый уровень. Часть 2» и «Линейная алгебра и геометрия. Продвинутый уровень. Часть 1», «Линейная алгебра и геометрия. Продвинутый уровень. Часть 2».

Цель изучения дисциплины (модуля): формирование у студентов глубокого теоретического понимания принципов обучения по данным, включая математические основы обобщающей способности, оценку сложности моделей и анализ ошибок обучения, а также развитие способности критически оценивать и выбирать модели на основе статистических критериев.

Задачи изучения дисциплины (модуля):

- освоить теоретические основы обобщающей способности: неравенства концентрации (например, Хёфдинга), понятие равномерной сходимости;
- познакомиться с мерами сложности моделей: размерность Вапника–Червоненкиса (VC-размерность), радемахеровская сложность;
- изучить фундаментальные теоремы теории обучения, включая теоремы о сходимости эмпирического риска (ERM) и оценки обобщения;
- проанализировать компромисс между смещением и дисперсией и его влияние на выбор модели;
- рассмотреть регуляризацию как инструмент управления сложностью модели (L1, L2, структурная регуляризация);
- понять принципы кросс-валидации и оценки обобщающей способности на практике;
- научиться интерпретировать поведение алгоритмов обучения с теоретической точки зрения.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- основные понятия теории обучения: модель, функция потерь, эмпирический риск, обобщающая способность, класс гипотез;

- базовые модели обучения (рас-обучение, онлайн-обучение) и их теоретические гарантии;

- ключевые меры сложности класса гипотез: vc -размерность и радемахеровскую сложность;

- фундаментальные теоремы обучения (границы обобщения, теоремы о сходимости, роль регуляризации);

- основные алгоритмы (svm, adaboost, ядерные методы, нейронные сети) с точки зрения их теоретических свойств;

уметь:

- формализовать задачу обучения в рамках статистической теории (пространство объектов, гипотез, функция потерь);

- вычислять или оценивать vc -размерность и радемахеровскую сложность для простых классов гипотез;

- применять метод минимизации эмпирического риска (erm) и структурной минимизации риска (sgm) для построения алгоритмов с гарантиями обобщения;

- использовать вероятностные неравенства (хёффдинга, чернова, ограниченных разностей) для получения границ обобщающей способности;

- реализовывать базовые алгоритмы статистического обучения (svm, adaboost) с помощью библиотек scikit-learn и интерпретировать результаты с точки зрения теории;

владеть:

- навыками построения и анализа математических моделей обучения с использованием аппарата теории вероятностей и математической статистики;

- методологией строгого анализа и оценки качества алгоритмов машинного обучения с точки зрения их теоретической обоснованности;

- методами получения оценок вычислительной сложности и требуемого объема выборки для различных моделей и алгоритмов;

- техникой применения вероятностных неравенств и методов концентрации для вывода теоретических гарантий обобщающей способности;

- пониманием текущих направлений исследований в области статистической теории обучения и умением интерпретировать результаты современных научных статей.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Код и содержание компетенции	Индикатор компетенции и перечень планируемых результатов обучения по дисциплине (модулю)
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Знает методы поиска и анализа информации в области разработки, основные принципы критической оценки источников информации и их релевантности.
	УК-1.2. Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем.
	УК-1.3. Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации.
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1. Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач.
	УК-2.2. Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения.
	УК-2.3. Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями.
ОПК-1. Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности	ОПК-1.1. Знает основные понятия, методы и результаты математического анализа, комплексного и функционального анализа, алгебры, аналитической и дифференциальной геометрии, топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов и теоретической механики.
	ОПК-1.2. Умеет применять фундаментальные математические знания и методы для анализа и решения профессиональных задач.
	ОПК-1.3. Имеет практический опыт выбора, разработки и реализации математических моделей и методов при решении задач профессиональной деятельности.
ОПК-6. Способен разрабатывать алгоритмы и компьютерные программы, пригодные для	ОПК-6.1. Знает принципы алгоритмизации, основные структуры данных, парадигмы программирования, методы разработки, тестирования и отладки компьютерных программ.
	ОПК-6.2. Умеет разрабатывать, документировать, тестировать и

практического применения	отлаживать алгоритмы и компьютерные программы, пригодные для практического применения, с использованием современных языков и сред разработки.
	ОПК-6.3. Имеет практический опыт разработки и сопровождения компьютерных программ и программных комплексов для решения прикладных задач.

3. Тематический план

№п/ п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Лекции	Практические занятия			
1	Основы и РАС-обучение	5	5		20	Домашнее задание Аудиторная работа
2	Неравенства и меры сложности	5	5	2	20	Домашнее задание Аудиторная работа Контрольная работа
3	SVM, ядра и ансамбли	5	5		20	Домашнее задание Аудиторная работа
4	Нейронные сети и регуляризация	5	5		20	Домашнее задание Аудиторная работа
5	Теория маржи и онлайн-обучение	5	5		20	Домашнее задание Аудиторная работа
6	Оценка качества и современные направления	2	2		10	Домашнее задание Аудиторная работа
7	Оценка качества и современные направления	3	5	2	10	Домашнее задание Аудиторная работа Контрольная работа
	Зачет с оценкой			4		Коллоквиум
Итого:		30	32	8	120	
Объем дисциплины (модуля) (в ак. ч.)		190				
Объем дисциплины (модуля) (в зач. ед.)		5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Основы и PAC-обучение	Постановка задачи. Риск и эмпирический риск PAC-обучение. Конечные классы гипотез
2	Неравенства и меры сложности	Хёффдинга, Чернова, ограниченных разностей VC-размерность. Примеры вычисления Радемахеровская сложность. Границы обобщения.
3	SVM, ядра и ансамбли	Метод опорных векторов (SVM) Ядерный трюк. RKHS Алгоритм AdaBoost
4	Нейронные сети и регуляризация	VC-размерность сетей. NTK. Структурная минимизация риска. L1, L2
5	Теория маржи и онлайн-обучение	Границы, основанные на марже Экспертные советы. Перцептрон
6	Оценка качества и современные направления	Кросс-валидация, доверительные интервалы для метрик Обучение с подкреплением. Открытые проблемы
7	Оценка качества и современные направления	От теории обобщения к практике. Сравнительный анализ подходов и современные направления. Контрольная работа 2

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Введение в статистическое обучение с примерами на языке R : учебник / Д. Гарет, Д. Уиттон, Т. Хасти, Р. Тибширани ; пер. с англ. С. Э. Мастицкого. - 2-е изд. испр. - Москва : ДМК Пресс, 2017. - 456 с. - ISBN 978-5-97060-495-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1873554>

2. Мастицкий, С. Э. Статистический анализ и визуализация данных с помощью R : практическое руководство / С. Э. Мастицкий, В. К. Шитиков. - 2-е изд. - Москва : ДМК Пресс, 2023. - 497 с. - ISBN 978-5-89818-601-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2108480>

Дополнительная литература:

1. Тимофеева, А. Ю. Вероятностные основы методов и алгоритмов анализа данных : учебное пособие / А. Ю. Тимофеева, Е. А. Хайленко. - Новосибирск : Изд-во НГТУ, 2020. - 82 с. - ISBN 978-5-7782-4166-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1866892>

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модуле), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное

Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Статистическая теория обучения» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары, контрольные работы, домашние задания, коллоквиум, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар — это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Аудиторная работа – активная работа студента на семинаре или лекции, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в аудиторной работе студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения, аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Контрольная работа – письменная работа с набором задач, которые нужно решить за ограниченное время.

Цель контрольной работы - получить специальные знания по одной или нескольким темам дисциплины (модуля) и продемонстрировать навыки их практического применения.

Коллоквиум – устные ответы на вопросы, список которых известен студенту заранее.

В процессе подготовки к коллоквиуму необходимо проанализировать учебные материалы, ознакомившись с лекциями, учебниками и дополнительными источниками, акцентируя внимание на ключевых темах. Рекомендуется создать структурированные конспекты, выделяя основные идеи, термины и формулы.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Статистическая теория обучения»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме **зачета с оценкой**, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	Зачтено	
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и
4	Удовлетворительно	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Статистическая теория обучения» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	25%	10	Набор задач по темам недели
Контрольные работы	35%	2	Письменная работа с набором задач, которые нужно решить за ограниченное время
Аудиторная работа	15%	15	Активная работа студента на семинаре или лекции
Зачет с оценкой	25%	1	Коллоквиум - устные ответы на вопросы, список которых известен студенту заранее

Формула расчёта итоговой оценки по дисциплине (модулю) «Статистическая теория обучения»: $0,25 \times \text{среднее за домашние задания} + 0,35 \times \text{среднее за контрольные работы} + 0,15 \times \text{среднее за аудиторную работу} + 0,25 \times \text{зачет с оценкой}$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

1. Даны выборка и функция потерь $\ell(y, f(x)) = (y - f(x))^2$. Запишите выражение для эмпирического риска. В чём разница между эмпирическим и истинным риском?
2. Приведите пример задачи, в которой высокая точность на обучающей выборке не гарантирует хорошего качества на тесте. Объясните, почему это происходит.
3. Постройте пример, иллюстрирующий компромисс между смещением и дисперсией при выборе степени полиномиальной регрессии.
4. Даны выборка и функция потерь $\ell(y, f(x)) = (y - f(x))^2$. Запишите выражение для эмпирического риска. В чём разница между эмпирическим и истинным риском?
5. Приведите пример задачи, в которой высокая точность на обучающей выборке не гарантирует хорошего качества на тесте. Объясните, почему это происходит.
6. Постройте пример, иллюстрирующий компромисс между смещением и дисперсией при выборе степени полиномиальной регрессии.
7. Даны выборка и функция потерь $\ell(y, f(x)) = (y - f(x))^2$. Запишите выражение для эмпирического риска. В чём разница между эмпирическим и истинным риском?
8. Приведите пример задачи, в которой высокая точность на обучающей выборке не гарантирует хорошего качества на тесте. Объясните, почему это происходит.
9. Постройте пример, иллюстрирующий компромисс между смещением и дисперсией при выборе степени полиномиальной регрессии.
10. Вычислите выборочную радемахеровскую сложность для множества из двух функций: $h_1(x)=1$, $h_2(x)=-1$, при фиксированной выборке $x_1, \dots, x_n$.

Примерные задания по контрольной работе

1. Дана модель бинарной классификации с функцией потерь 0-1. Запишите выражения для истинного и эмпирического рисков. В чём заключается проблема минимизации эмпирического риска?
2. Что такое PAC-обучаемость? Сформулируйте определение в терминах ϵ и δ .
3. Найдите VC-размерность класса всех прямоугольников на плоскости, ориентированных по осям координат. Приведите краткое обоснование.
4. С помощью неравенства Хёфдинга оцените минимальный объём выборки n , необходимый для того, чтобы с вероятностью не менее 0.95 эмпирический риск отличался от истинного не более чем на 0.05.
5. Объясните, как радемахеровская сложность используется для получения неравенства обобщающей способности. Запишите общую форму границы.
6. Какой вклад вносит регуляризация в компромисс между смещением и дисперсией? Приведите пример из линейной регрессии.
7. В чём заключается идея структурной минимизации риска (SRM)? Как она реализуется в методе опорных векторов?
8. Пусть H — конечный класс гипотез, $|H|=50$. Получите верхнюю оценку вероятности того, что для некоторой гипотезы $h \in H$ разница между истинным и эмпирическим рисками превысит ϵ .

Примерные задания для семинаров

1. Постройте пример, когда минимизация эмпирического риска приводит к переобучению.
2. Найдите VC-размерность класса всех кругов на плоскости.
3. Постройте множество из 3 точек, разбиваемое всеми способами с помощью пороговых функций на прямой.

4. Выведите границу обобщающей способности для конечного класса гипотез с помощью неравенства объединения и Хёфдинга.
5. Обсудите, почему равномерная сходимость важна для теории обучения.
6. Сравните границы, основанные на VC-размерности и радемахеровской сложности.
7. Покажите, как L2-регуляризация влияет на решение в линейной регрессии.
8. Нарисуйте график зависимости обобщающей ошибки от сложности модели.
9. Обозначьте участки переобучения и недообучения.
10. Обсудите: как кросс-валидация помогает выбрать оптимальный уровень регуляризации?

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция	Индикатор
1.	Дайте определение эмпирического риска (одним предложением).	Эмпирический риск — это среднее значение функции потерь на обучающей выборке	УК-1	УК-1.1
2.	Назовите основную цель статистической теории обучения (одно слово).	Обобщающая способность	ОПК-1	ОПК-1.1
3.	Запишите формулу эмпирического риска для выборки $(x_i, y_i)_{i=1}^n$ и функции потерь ℓ .	$R_{\text{emp}}(f) = (1/n) \sum_{i=1}^n \ell(y_i, f(x_i))$	ОПК-1	ОПК-1.2
4.	Укажите, что минимизируется в методе ERM (одно слово).	Эмпирический риск	ОПК-1	ОПК-1.1
5.	Дайте определение PAC-обучаемости (кратко).	Класс гипотез PAC-обучаем, если для любых $\epsilon > 0, \delta > 0$ существует алгоритм, гарантирующий ошибку $\leq \epsilon$ с вероятностью $\geq 1 - \delta$	ОПК-1	ОПК-1.1
6.	Назовите тип гарантии, предоставляемой PAC-обучением (одно слово).	Вероятностная	УК-1	УК-1.2
7.	Запишите неравенство Хёфдинга для суммы независимых величин $Z_i \in [0][1]$.	$P(\sum Z_i - E[\sum Z_i] \geq t) \leq 2 \exp(-2t^2/n)$	ОПК-1	ОПК-1.2
8.	Назовите меру сложности класса гипотез, связанную с числом способов разметки точек.	VC-размерность	ОПК-1	ОПК-1.1
9.	Дайте определение VC-размерности (кратко).	Максимальное число точек, которые можно разделить всеми возможными способами с помощью функций из данного класса	ОПК-1	ОПК-1.1

10.	Найдите VC-размерность класса пороговых функций на прямой (одно число).	1	ОПК-1	ОПК-1.2
11.	Укажите VC-размерность линейных классификаторов в R^d (одно выражение).	$d+1$	ОПК-1	ОПК-1.2
12.	Назовите меру сложности, основанную на случайных шумах Радемахера.	Радемахеровская сложность	ОПК-1	ОПК-1.1
13.	Запишите выражение для выборочной радемахеровской сложности класса H .	$\hat{R}_n(H) = E_{\sigma}[\sup_{h \in H} (1/n) \sum_{i=1}^n \sigma_i h(x_i)], \sigma_i = \pm 1$	ОПК-1	ОПК-1.2
14.	Назовите алгоритм ансамблевого обучения, основанный на взвешивании ошибочных объектов.	AdaBoost	ОПК-6	ОПК-6.1
15.	Укажите, как регуляризация L2 влияет на веса модели (одно слово).	Уменьшает	ОПК-1	ОПК-1.2
16.	Запишите общий вид структурированного риска (SRM).	$R_{\text{struct}}(f) = R_{\text{emp}}(f) + \lambda \cdot \text{complexity}(f)$	ОПК-1	ОПК-1.2
17.	Назовите компромисс между смещением и дисперсией (англ.).	Bias-variance tradeoff	УК-2	УК-2.2
18.	Укажите метод оценки обобщающей способности с разбиением данных на k частей.	Кросс-валидация	ОПК-6	ОПК-6.2
19.	Назовите алгоритм онлайн-обучения для линейной классификации (одно слово).	Перцептрон	ОПК-1	ОПК-1.1
20.	Укажите один критерий остановки для итеративных алгоритмов обучения (одно слово).	Сходимость	УК-2	УК-2.2