

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
**Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Нижегородский государственный университет  
им. Н.И. Лобачевского»**

Институт информационных технологий, математики и механики

---

УТВЕРЖДЕНО  
президиумом Ученого совета ННГУ  
протокол от  
«30» ноября 2022 г. № 13

**Рабочая программа дисциплины**

Теория машинного обучения

---

Уровень высшего образования  
магистратура

---

Направление подготовки  
02.04.02 «Фундаментальная информатика и информационные технологии»

---

Направленность образовательной программы  
**Искусственный интеллект**

---

Форма обучения  
очная

---

Нижний Новгород  
2023 год

## 1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.06.02 «Теория машинного обучения» относится к дисциплинам по выбору части Блока 1, формируемой участниками образовательных отношений, «Дисциплины (модули)» направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии», направленность «Искусственный интеллект». Дисциплина преподается во 3 семестре.

№ Варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
1	Блок 1. Дисциплины (модули) Часть, формируемая участниками образовательных отношений. Дисциплина по выбору	Дисциплина Б1.В.ДВ.06.02 «Теория машинного обучения» относится к части ООП направления подготовки 02.04.02 Фундаментальная информатика и информационные технологии, формируемой участниками образовательных отношений.

## 2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

3.

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции* (код, содержание индикатора)	Результаты обучения по дисциплине**	
ПК-8. Способен к разработке новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности.	ПК-8.1. Знать основы разработки новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности	<u>Знать</u> теоретические основы машинного обучения: постановки задач обучения по прецедентам, обучения без учителя; основные алгоритмы решения задачи восстановления регрессии (метод наименьших квадратов, гребневая регрессия, «лассо», регрессия методом главных компонент, частичные наименьшие квадраты); основные алгоритмы решения задачи классификации (байесов классификатор, алгоритм ближайших соседей, линейная регрессия, линейный и квадратичный дискриминантный анализ, логистическая регрессия, Персептрон Розенблатта, алгоритм	<i>Собеседование, контрольная работа</i>

		<p>обратного распространения ошибки для обучения нейронной сети, машина опорных векторов, деревья решений, градиентный бустинг, баггинг, случайные деревья)</p> <p><u>Владеть</u> современными инструментальными вычислительными средствами теории машинного обучения.</p>	
	<p>ПК-8.2.</p> <p>Иметь навыки разработки новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности</p>	<p><u>Знать</u> методы практического применения теоретических знаний в области фундаментальной информатики и информационных технологий</p> <p><u>Уметь</u> профессионально разрабатывать и проводить процедуры тестирования алгоритмов кластеризации (методы из теории графов, методы центров тяжести и медиан, EM-метод).</p> <p><u>Владеть</u> методами практического применения теоретических знаний в области фундаментальной информатики и информационных технологий.</p>	<p><i>Контрольная работа</i></p>
	<p>ПК-8.3.</p> <p>Иметь навыки управления разработкой и развитием новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности</p>	<p><u>Знать</u> подходы к решению задач иерархической кластеризации; экспериментальные методы оценки качества обучения; основы теории Вапника–Червоненкиса.</p> <p><u>Уметь</u> использовать имеющиеся знания для решения практических задач машинного обучения; оценивать на практике качество обучения используемых моделей.</p> <p><u>Владеть навыками</u> решения практических задач с использованием методов машинного обучения; использования среды статистических вычислений R или библиотеки ScikitLearn для решения задач машинного обучения.</p>	<p><i>Собеседование</i></p>

### 3. Структура и содержание дисциплины

#### 3.1. Трудоемкость дисциплины

<b>Общая трудоемкость</b>	<b>3 ЗЕТ</b>
<b>Часов по учебному плану</b>	<b>108</b>
<b>в том числе:</b>	
<b>аудиторные занятия (контактная работа):</b>	<b>33</b>
- занятия лекционного типа	16
- занятия семинарского типа	16
занятия лабораторного типа	
- текущий контроль (КСР)	1
<b>самостоятельная работа</b>	<b>75</b>
<b>Промежуточная аттестация</b>	<b>зачет</b>

#### 3.2 Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины,  форма промежуточной аттестации по дисциплине	Всего (часы)	в том числе				
		контактная работа (работа во взаимодействии с преподавателем), часы, из них				Самостоятельная работа студента, часы
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего контактных часов	
<b>1. Введение.</b> Основные задачи теории обучения.	4	1		0	1	3
<b>2. Основы теории Вапника–Червоненкиса.</b> Лемма Бернштейна–Чернова. Равномерная сходимости эмпирического риска к ожидаемому риску в конечном случае. Размерность Вапника–Червоненкиса. Примеры: пороговые функции, нейронные сети, деревья решений. $\epsilon$ -сеть. Теорема Вапника–Червоненкиса о равномерной сходимости эмпирического риска к ожидаемому в бесконечном случае.	25	3		4	7	18
<b>3. Основы теории PAC-обучения.</b> Определение «вероятно почти корректного» (PAC–probably approximately correct learning) обучения. Обучение конъюнкциям. Труднорешаемость задачи обучения	26	4		4	8	18

3-ДНФ формулам. Обучение 3-КНФ формулам.						
<b>4. Лезвие Оккама.</b> Использования принципа лезвия Оккама при обучении конъюнкциям. Обучение деревьям решений.	26	4		4	8	18
<b>5. Точное обучение.</b> Обучение с помощью вопросов. Вопросы принадлежности и эквивалентности. Размерность научения. Верхние и нижние оценки сложности обучения на основе размерности научения.	26	4		4	8	18
Текущий контроль (КСР)	1				1	
<b>Итого</b>	<b>108</b>	<b>16</b>		<b>16</b>	<b>33</b>	<b>75</b>

Практические занятия (лабораторные занятия) организуются, в том числе в форме практической подготовки, которая предусматривает участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

Практическая подготовка предусматривает: вычисление размерности Вапника–Червоненкиса класса пороговых функций; вычисление размерности Вапника–Червоненкиса класса нейронных сетей; вычисление размерности Вапника–Червоненкиса класса деревьев решений; использование принципа лезвия Оккама при обучении.

На проведение практических занятий (лабораторных работ) в форме практической подготовки отводится 16 часов.

Практическая подготовка направлена на формирование и развитие:

- практических навыков в соответствии с профилем ОП: Разработка, тестирование, оптимизация программного обеспечения (ПО). Разработка технической документации на продукцию в сфере ИТ.
- компетенций – ПК-8: Способен к разработке новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности (ПК-8.3: Имеет практический опыт анализа и интерпретации сложных информационных систем).

Текущий контроль успеваемости реализуется в рамках занятий лабораторного типа, групповых или индивидуальных консультаций. Промежуточная аттестация проходит в традиционных формах (зачет).

#### **4. Учебно-методическое обеспечение самостоятельной работы обучающихся**

Самостоятельная работа обучающихся осуществляется в виде работы с рекомендованной обязательной и дополнительной литературой, подготовке к лекциям, подготовке к экзамену и выполнения лабораторных работ. Контрольные вопросы и задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 6.2.

**5. Фонд оценочных средств для промежуточной аттестации по дисциплине, включающий:**

**5.1. Описание шкал оценивания результатов обучения по дисциплине**

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности компетенций						
	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	Не зачтено		зачтено				
<u>Знания</u>	Отсутствие знаний теоретического материала.  Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки.	Минимально допустимый уровень знаний. Допущено много негрубых ошибок.	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько незначительных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений . Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения.  Имели место грубые ошибки.	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания но не в полном объеме.	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи . Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения, решены все основные задачи с отдельными незначительными недочетами, выполнены все задания в полном объеме.	Продemonстрированы все основные умения,. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие владения материалом. Невозможность оценить наличие навыков вследствие отказа	При решении стандартных задач не продемонстрированы базовые навыки.  Имели место грубые	Имеется минимальный набор навыков для решения стандартных задач с некоторыми	Продemonстрированы базовые навыки  при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки  при решении стандартных задач без недочетов.	Продemonстрированы навыки  при решении нестандартных задач без ошибок и недочетов.	Продemonстрирован творческий подход к решению нестандартных задач

	обучающего от ответа	ошибки.	недочетами				
--	-------------------------	---------	------------	--	--	--	--

### Шкала оценки при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	Превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
	Отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	Очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	Хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «хорошо»
	Удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	Неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

## 5.2 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения по ПК-8

### 5.2.1 Контрольные вопросы

	Вопрос	Код формируемой компетенции
1.	Лемма Бернштейна–Чернова.	ПК-8
2.	Равномерная сходимость эмпирического риска к ожидаемому риску в конечном случае.	ПК-8
3.	Размерность Вапника–Червоненкиса класса пороговых функций.	ПК-8
4.	Размерность Вапника–Червоненкиса класса нейронных сетей.	ПК-8
5.	Размерность Вапника–Червоненкиса для класса деревьев решений.	ПК-8
6.	$\varepsilon$ -сеть. Теорема Вапника–Червоненкиса о равномерной сходимости эмпирического риска к ожидаемому в бесконечном случае.	ПК-8
7.	Определение «вероятно почти корректного» (РАС–probably approximately correct learning) обучения.	ПК-8
8.	РАС-обучение. Обучение конъюнкциям.	ПК-8
9.	РАС-обучение. Труднорешаемость задачи обучения 3-ДНФ формулам.	ПК-8
10.	РАС-обучение. Обучение 3-КНФ формулам.	ПК-8
11.	Лезвие Оккама. Использование принципа лезвия Оккама при обучении конъюнкциям.	ПК-8
12.	Использование принципа лезвия Оккама при обучении деревьям решений.	ПК-8
13.	Обучение с помощью вопросов. Вопросы принадлежности и эквивалентности. Размерность научения.	ПК-8
14.	Верхние оценки сложности обучения с помощью вопросов принадлежности на основе размерности научения.	ПК-8
15.	Нижние оценки сложности обучения с помощью вопросов на основе размерности научения.	ПК-8

### 5.2.2. Типовые тестовые задания для оценки сформированности компетенции ПК-8

1. Функцию  $f : \mathbf{R}^d \rightarrow \{0, 1\}$  назовем *ящиком*, если существуют вещественные числа  $a_1, a_2, \dots, a_d, b_1, b_2, \dots, b_d$ , такие, что  $f(x) = 1$  тогда и только тогда, когда  $a_i \leq x \leq b_i$  ( $i = 1, 2, \dots, d$ ).  
Найти функцию роста и размерность Вапника–Червоненкиса для класса всех ящиков.  
Проиллюстрировать на этом примере лемму Зауэра.
2. Пусть  $T_h$  — множество всех функций  $f : \mathbf{R}^d \rightarrow \{0, 1\}$ , вычисляемых бинарными деревьями решений, высоты не выше  $h$ . Найти функцию роста и размерность Вапника–Червоненкиса для класса  $T_h$ . Проиллюстрировать на этом примере лемму Зауэра.



3. Пусть  $H_d$  — множество всех булевых функций  $f : \{0, 1\}^d \rightarrow \{0, 1\}$ , представимых ДНФ, в которых каждая элементарная конъюнкция представляет собой одиночный символ, обозначающий переменную (без отрицания). Найти функцию роста и размерность Вапника–Червоненкиса для класса  $H_d$ . Проиллюстрировать на этом примере лемму Зауэра.
4. Функцию  $f : \mathbf{R}^2 \rightarrow \{0, 1\}$  назовем полигоном (точнее:  $k$ -вершинным полигоном), если найдется выпуклый  $k$ -угольник  $M$ , такой, что  $f(x) = 1$  тогда и только тогда, когда  $x$  принадлежит  $M$ . Пусть  $P$  — множество всех полигонов, а  $P_k$  — множество всех  $k$ -вершинных полигонов. Найти  $VC P$  и  $VC P_k$ .
5. Привести пример бесконечного класса  $\mathcal{F}$ , для которого  $VC \mathcal{F} = 1$ .
6. *Размерность Вапника–Червоненкиса для задачи восстановления регрессии.* Пусть  $\mathcal{F}$  — некоторый класс функций  $f : \mathcal{X} \rightarrow \mathcal{Y}$ . Размерностью Вапника–Червоненкиса для класса  $\mathcal{F}$  называется  $VC \mathcal{F}'$ , где

$$\mathcal{F}' = \{I(f(x) - y) : f \in \mathcal{F}, y \in \mathcal{Y}\}.$$

Найти размерность Вапника–Червоненкиса для класса  $\{\sin \alpha x : \alpha \in \mathbf{R}\}$ .

7. Пусть  $\mathcal{X}$  — подмножество в  $\mathbf{R}^d$ , а  $\mathcal{F}$  — некоторое множество функций, отображающих  $\mathcal{X}$  в  $\{0, 1\}$ . Введем класс  $\mathcal{F}' = \{f \vee g : f, g \in \mathcal{F}\}$ , состоящий из дизъюнкций каждой пары функций в  $\mathcal{F}$ . Доказать, что для функции роста класса  $\mathcal{F}'$  справедливо неравенство  $\Delta(\mathcal{F}', N) \leq \Delta(\mathcal{F}, N)^2$ . Воспользовавшись леммой Зауэра, доказать, что  $VC \mathcal{F}' \leq 10 VC \mathcal{F}$ . Что изменится, если вместо дизъюнкций рассмотреть (а) все конъюнкции, (б) суммы по модулю 2?

## 6. Учебно-методическое и информационное обеспечение дисциплины

### а) Основная литература:

1. Воронцов К.В. Машинное обучение. Курс лекций.  
<http://www.machinelearning.ru>.  
<http://www.intuit.ru/studies/courses/13844/1241/info>
2. Золотых Н.Ю. Машинное обучение. Курс лекций. Нижний Новгород: ННГУ, 2007.  
<http://www.uic.unn.ru/~zny/ml>

### б) Дополнительная литература

1. Введение в аналитику больших массивов данных  
<http://www.intuit.ru/studies/courses/12385/1181/info>

### в) программное обеспечение и Интернет-ресурсы

- 1) The R Project for Statistical Computing URL: <https://www.r-project>
- 2) Welcome to Python.org URL: <https://www.python.org>
- 3) scikit-learn: machine learning in Python URL: <http://scikit-learn.org>

## **7. Материально-техническое обеспечение дисциплины (модуля)**

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой, оснащенные оборудованием и техническими средствами обучения: компьютерный класс, проектор, экран.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Учебная и научная литература, учебно-методические материалы, представленные в библиотечном фонде, в электронных библиотеках и на кафедре математического обеспечения и суперкомпьютерных технологий.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению 02.04.02 «Фундаментальная информатика и информационные технологии».

Автор д.ф.-м.н., проф. Н. Ю. Золотых

Заведующий кафедрой АГиДМ Н.Ю. Золотых

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики от 30 ноября 2022 года, протокол № 3.