

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
решением Ученого совета ННГУ
протокол от 30 ноября 2022 г. № 13

Рабочая программа дисциплины

Машинное обучение

Уровень высшего образования
магистратура

Направление подготовки
010402 Прикладная математика и информатика

Направленность образовательной программы
Анализ данных в прикладных областях

Форма обучения
очная

Нижегород
2023 год

1. Место и цели дисциплины в структуре ООП

Место дисциплины в структуре ООП

№ варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
1	Блок 1. Дисциплины (модули) Обязательная часть	Дисциплина Б1.О.08, «Машинное обучение» относится к обязательной части ООП направления подготовки "01.04.02 Прикладная математика и информатика".

Целями освоения дисциплины являются формирование у студентов компетенций в соответствии с требованиями ОС ННГУ по направлению подготовки 01.04.02 – Прикладная математика и информатика. Содержание дисциплины направлено на освоение алгоритмов и методов машинного обучения (machine learning) основ статистической теории Вапника–Червоненкиса; формирование умений и навыков в решении практических задач с использованием методов машинного обучения.

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

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	
ОПК-1. Способен решать актуальные задачи фундаментальной и прикладной математики	ОПК-1.1. Знает современные методы решения задач фундаментальной и прикладной математики	ЗНАТЬ: основные алгоритмы решения задачи восстановления регрессии и классификации, алгоритмы кластеризации	Собеседование (экзамен)
	ОПК-1.2. Умеет использовать фундаментальные знания и практический опыт в профессиональной деятельности	УМЕТЬ: использовать имеющиеся знания для решения практических задач машинного обучения	Контрольная работа (текущий контроль)
	ОПК-1.3. Имеет навыки решения актуальных и значимых проблем фундаментальной и прикладной математики	ИМЕТЬ НАВЫКИ: использования среды статистических вычислений R или библиотеки ScikitLearn для решения задач машинного обучения	Контрольные работы
ОПК-3. Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	ОПК-3.1. Знает современные методы анализа математических моделей при решении задач в области профессиональной деятельности	ЗНАТЬ основные концептуальные и теоретические модели машинного обучения	Собеседование (экзамен)
	ОПК-3.2. Умеет разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	УМЕТЬ разрабатывать и анализировать концептуальные и теоретические модели машинного обучения	Контрольная работа (текущий контроль)

	ОПК-3.3. Имеет навыки разработки новых математических моделей при решении задач в области профессиональной деятельности	ИМЕТЬ НАВЫКИ: оценивать на практике качество обучения используемых моделей	Контрольные работы
ПК-1. Способен разрабатывать и анализировать концептуальные и теоретические модели решаемых проблем и задач в области профессиональной деятельности	ПК-1.1. Знает методы разработки и анализа концептуальных и теоретических моделей решаемых проблем и задач в области профессиональной деятельности.	ЗНАТЬ основные алгоритмы и методы машинного обучения основы языка Python или среды вычислений R	Собеседование
	ПК-1.2. Умеет применять методы разработки и анализа концептуальных и теоретических моделей решаемых проблем и задач в области профессиональной деятельности	УМЕТЬ использовать методы машинного обучения на практике, умения оценивать качество методов работать с библиотекой Scikit-Learn или средой для статистических вычислений R	Контрольная работа (текущий контроль)
	ПК-1.3. Имеет опыт применения методов разработки и анализа концептуальных и теоретических моделей решаемых проблем и задач в области профессиональной деятельности	ИМЕТЬ НАВЫКИ: решения практических задач с использованием методов машинного обучения	Контрольные работы

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

	Очная форма обучения
Общая трудоемкость	7 ЗЕТ
Часов по учебному плану	252
в том числе	
аудиторные занятия (контактная работа):	66
- занятия лекционного типа	32
- занятия семинарского типа	32
- занятия лабораторного типа	
- текущий контроль (КСР)	2
самостоятельная работа	150
Промежуточная аттестация – экзамен	36

3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины	Всего (часы)	В том числе				
		Контактная работа (работа во взаимодействии с преподавателем), часы. Из них				Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего	
1. Содержательная постановка задачи машинного обучения	6	2	2		4	2

2. Вероятностная постановка задачи обучения по прецедентам	6	2	2		4	2
3. Экспериментальные методы оценки качества обучения	6	2	2		4	2
4. Метод наименьших квадратов для решения задачи восстановления регрессии	7	2	2		4	3
5. Проблема переобучения при решении задачи восстановления регрессии	7	2	2		4	3
6. Метод ближайших соседей для решения задачи классификации	7	2	2		4	3
7. Наивный байесовский классификатор	7	2	2		4	3
8. Линейный дискриминантный анализ	7	2	2		4	3
9. Нейронные сети. Персептрон Розенблатта	7	2	2		4	3
10. Машина опорных векторов	7	2	2		4	3
11. Деревья решений	7	2	2		4	3
12. Ансамбли решающих правил	7	2	2		4	3
13. Обучение без учителя	7	2	2		4	3
14. Иерархическая кластеризация	7	2	2		4	3
15. Основы теории Вапника–Червоненкиса	11	4	4		8	3
Текущий контроль (КСР)	2				2	
Промежуточная аттестация – экзамен	36					
Итого	144	32	32	0	66	42

Текущий контроль успеваемости реализуется в формах опросов на занятиях семинарского типа. Промежуточная аттестация проходит в традиционной форме (экзамен).

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Ниже приводятся виды самостоятельной работы студентов, порядок их выполнения и контроля, приводится учебно-методическое обеспечение самостоятельной работы по ее отдельным видам и разделам дисциплины.

Виды самостоятельной работы студентов:

- проработка теоретического материала лекционных занятий;
- подготовка домашних заданий к научно-практическим занятиям;
- подготовка к выполнению письменных контрольных работ;
- подготовка к промежуточной аттестации в форме зачета или экзамена

4.1. Проработка теоретического материала лекционных занятий

Выполняется самостоятельно с использованием лекционных материалов и материалов, разобранных в литературе (список обязательной и дополнительной литературы приводится).

Контроль выполняется в форме проведения ежемесячного письменного экспресс - опроса по понятиям, фактам, формулировкам, выполняемого в течение 15 минут на практических занятиях. Экспресс – опрос оценивается оценками «Зачтено» – «Не зачтено».

4.2. Подготовка домашних заданий к научно-практическим занятиям

Домашние задания выдаются на практических занятиях.

Проверка выполнения домашних заданий проводится в начале каждого практического занятия. Используется две формы контроля: – выборочная проверка выполнения заданий у двух-трех человек из группы; – проверка в форме коллективного обсуждения у доски результатов выполнения отдельных заданий одним или двумя студентами.

4.3. Подготовка к выполнению письменных контрольных работ

В течение учебного семестра проводится две домашние контрольные работы по материалам всех разделов лекционного курса.

Для подготовки к контрольным работам рекомендуется повторно прочитать теоретические разделы в задачнике, просмотреть полезные разделы в соответствующих

источниках из списка рекомендованной литературы, а также самостоятельно решать несколько задач по теме контрольной работы из указанного задачника.

4.4. Подготовка к промежуточной аттестации в форме экзамена

В качестве методических материалов при подготовке к экзамену рекомендуется использовать собственные конспекты лекций и источники, рекомендованные в списке литературы раздела 7.

Контрольные вопросы и задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

5. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю), 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 несущественных 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 несущественными 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 обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки.	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами.	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов.	Продemonстрирован творческий подход к решению нестандартных задач.

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

5.2.1. Типовые тестовые задания для оценки сформированности компетенции ОПК-3, ПК-1

- Какая из следующих задач является задачей обучения без учителя?
 - восстановление регрессии
 - классификация
 - кластеризация (+)
- Какая из следующих задач является задачей обучения с учителем?
 - классификация (+)
 - кластеризация
 - восстановление функции распределения
- Переобучением называется
 - явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много больше ошибки на обучающей выборке (+)
 - явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много меньше ошибки на обучающей выборке
 - явление, возникающее при решении задач обучения с учителем, при котором решающее правило допускает большую ошибку на тестовой выборке

5.2.2. Типовые задания для оценки сформированности компетенции ОПК-1, ПК-1

- Дана обучающая выборка

x_1	0	1	1	0	0	1	1	2	6
x_2	3	3	1	0	1	1	2	3	1
y	0	0	0	0	1	1	1	1	1

Методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравнение разделяющей поверхности.

- Дана обучающая выборка (см. таблицу выше). Методом квадратичного дискриминантного анализа построить дискриминантные функции.
- Дана обучающая выборка (см. таблицу выше). С помощью наивного байесова классификатора оценить вероятности $P(Y = 1 | x_1 = 1, x_1 = 2)$
- Дана обучающая выборка:

x	-1	0	0	1	2
y	1	-2	1	7	8

- Методом наименьших квадратов построить полиномиальную модель вида $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$.
- Дана обучающая выборка (см. выше). Методом ридж-регрессии построить полиномиальную модель вида $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, если параметр регуляризации $\lambda = 2$.

5.2.3. Контрольные вопросы

Вопрос	Код компетенции
1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов лассификатор.	ОПК-1
2. Экспериментальные методы оценки качества обучения.	ОПК-1
3. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков).	ОПК-1
4. Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.	ПК-1
5. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.	ОПК-3
6. Наивный байесовский классификатор.	ПК-1
7. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.	ОПК-1
8. Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях.	ОПК-3
9. Машина опорных векторов. Ядра и спрямляющие пространства.	ПК-1
10. Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений.	ПК-1
11. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосование. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (случайный лес.).	ОПК-3
12. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких	ПК-1

множеств. ЕМ-алгоритм.	
13. Иерархическая кластеризация. Агломеративные и разделяющие методы.	ПК-1
14. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска.	ОПК-3
15. Размерность Вапника–Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.	ОПК-1

6. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература:

1. Золотых Н.Ю. Машинное обучение. Курс лекций. Нижний Новгород: ННГУ, 2007.

<http://www.uic.nnov.ru/~zny/ml>

2. Воронцов К.В. Машинное обучение. Курс лекций. <http://www.machinelearning.ru>.

б) дополнительная литература:

3. Курс «Машинное обучение» – <https://www.intuit.ru/studies/courses/13844/1241/info>.

в) программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. The R Project for Statistical Computing <https://www.r-project>

2. Welcome to Python.org <https://www.python.org/>

3. scikit-learn: machine learning in Python scikit-learn.org/

7. Материально-техническое обеспечение дисциплины

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой (лекционного и семинарского типа), оснащенные оборудованием и техническими средствами обучения.

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

Программа составлена в соответствии с требованиями ОС ННГУ по направлению 01.04.02 Прикладная математика и информатика.

Автор: д.ф.-м.н., доц. _____ Н.Ю. Золотых

Рецензент: _____

Зав. кафедрой: д.ф.м.н., доц. _____ Н.Ю. Золотых

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