

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение высшего
образования**

«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 (специальность)

020402 Фундаментальная информатика и информационные технологии

Профиль подготовки (специализация)

Вероятностное моделирование и анализ данных

Форма обучения

очная

Нижний Новгород

2023

1. Место дисциплины в структуре ООП.

Дисциплина Б1.О.09 «Машинное обучение» относится к Части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)» направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии» профиля подготовки «Вероятностное моделирование и анализ данных». Дисциплина преподается в 3 семестре. Трудоемкость дисциплины составляет 6 зачетных единицы, 216 час., экзамен.

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**	
ПК-3. Способен эксплуатировать, разрабатывать и развивать системное и прикладное программное обеспечение, новые информационные технологии на основе анализа современного состояния науки и информационных технологий, и управлять такой эксплуатацией и разработкой в области профессиональной деятельности	ПК-3.1. Знать проблематику и методы научных исследований и опытно-конструкторских ИТ-разработок в области когнитивных систем (КС).	Знать алгоритмы и методы машинного обучения	собеседование
	ПК-3.2. Иметь навыки выполнения научных исследований и опытно-конструкторских ИТ-разработок в области КС	Уметь использовать методы и алгоритмы машинного обучения на практике, оценивать качество методов. Владеть современными программными средствами для решения задач машинного обучения: библиотекой Scikit-Learn, или средой языка Python или средой R для статистический вычислений; опытом реализации программных систем для решения практических задач с использованием методов машинного обучения	собеседование, задания

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

Общая трудоемкость	6 ЗЕТ
Часов по учебному плану	216
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	114
Промежуточная аттестация - экзамен	36

3.2 Содержание дисциплины

№ п/п	Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Часов					
		Всего	В том числе				Самостоятельная работа обучающегося
			Контактная работа (работа во взаимодействии с преподавателем), часы из них				
			Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего	
1.	Содержательная постановка задачи машинного обучения. Дедуктивное и индуктивное обучения. Признаковое описание объектов. Обучения с учителем (обучение по прецедентам). Решающая функция (решающее правило). Обобщающая способность решающей функции (проблема качества обучения). Задачи классификации и задачи восстановления регрессии. Обучение без учителя. Примеры практических задач.	8	2	2		4	7
2.	Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Регрессионная функция. Байесов классификатор. Метод ближайшего соседа в задачах классификации и восстановления регрессии.	9	2	2		4	8
3.	Экспериментальные методы оценки качества обучения. Разделение данных на обучающую и тестовую выборки. Метод скользящего контроля.	9	2	2		4	8
4.	Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная	9	2	2		4	8

	регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков).						
5.	Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.	9	2	2		4	8
6.	Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.	9	2	2		4	8
7.	Наивный байесовский классификатор.	9	2	2		4	8
8.	Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.	8	2	2		4	7
9.	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.	8	2	2		4	7
10.	Машина опорных векторов. Ядра и спрямляющие пространства.	10	2	2		4	8
11.	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.	10	2	2		4	7
12.	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.).	10	2	2		4	8
13.	Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. EM-алгоритм.	10	2	2		4	7

14.	Иерархическая кластеризация. Агломеративные и разделяющие методы.	10	2	2		4	8
15	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. Размерность Вапника–Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.	14	4	4		8	7
	Текущий контроль	2		2			
	Промежуточная аттестация: экзамен	36					
	Итого	144	32	34		66	114

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

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся осуществляется в виде работы с рекомендованной обязательной и дополнительной литературой (приведена в разделе 6), 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. 5.2.

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 теоретическо	Уровень знаний ниже минимальны	Минимально допустимый уровень	Уровень знаний в объеме,	Уровень знаний в объеме,	Уровень знаний в объеме,	

	го материала. Невозможнос ть оценить полноту знаний вследствие отказа обучающего я от ответа	х требований. Имели место грубые ошибки.	знаний. Допущено много негрубых ошибки.	соответствующ ем программе подготовки. Допущено несколько негрубых ошибок	соответствующ ем программе подготовки. Допущено несколько несущественн ых ошибок	соответствую щем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающе м программу подготовки.
<u>Умения</u>	Отсутствие минимальны х умений . Невозможнос ть оценить наличие умений вследствие отказа обучающего я от ответа	При решении стандартных задач не продемонстр ированы основные умения. Имели место грубые ошибки.	Продемонстр ированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания но не в полном объеме.	Продемонстри рованы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продемонстри рованы все основные умения. Решены все основные задачи . Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продемонстр ированы все основные умения, реше ны все основные задачи с отдельными несуществен ным недочетами, выполнены все задания в полном объеме.	Продемонстр ированы все основные умения,. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие владения материалом. Невозможнос ть оценить наличие навыков вследствие отказа обучающего я от ответа	При решении стандартных задач не продемонстр ированы базовые навыки. Имели место грубые ошибки.	Имеется минимальны й набор навыков для решения стандартных задач с некоторыми недочетами	Продемонстри рованы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстри рованы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продемонстр ированы навыки при решении нестандартн ых задач без ошибок и недочетов.	Продемонстр ирован творческий подход к решению нестандартн ых задач

Шкала оценивания при собеседовании

Результат ответов	Оценка
Студент дает верные развернутые ответы на вопросы преподавателя	зачтено
Студент отвечает только на часть вопросов, или отвечает с ошибками, или не дает развернутого ответа на вопросы	не зачтено

Шкала оценки при промежуточной аттестации

Оценка	Уровень подготовки
--------	--------------------

	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»
зачтено	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 Контрольные вопросы

Вопрос	Код формируемой компетенции
1) Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор.	ПК-3.1
2) Экспериментальные методы оценки качества обучения.	ПК-3.1
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).	ПК-3.1

4) Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.	ПК-3.1
5) Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.	ПК-3.1
6) Наивный байесовский классификатор.	ПК-3.1
7) Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.	ПК-3.1
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.	ПК-3.1
9) Машина опорных векторов. Ядра и спрямляющие пространства.	ПК-3.1
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-3.1
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.).	ПК-3.1
12) Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. EM-алгоритм.	ПК-3.1
13) Иерархическая кластеризация. Агломеративные и разделяющие методы.	ПК-3.1
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.	ПК-3.1
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.	ПК-3.1

5.2.2. Типовые задания/задачи для оценки сформированности компетенции ПК-10.2

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.

2. Дана обучающая выборка (см. таблицу выше). Методом квадратичного дискриминантного анализа построить дискриминантные функции.

3. Дана обучающая выборка (см. таблицу выше). С помощью наивного байесова классификатора оценить вероятности $P(Y = 1 | x_1 = 1, x_1 = 2)$
4. Дана обучающая выборка:

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

Методом наименьших квадратов построить полиномиальную модель вида $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$.

5. 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. Типовые вопросы для собеседования для оценки компетенции ПК-3

1. Что такое эмпирический риск?
2. Продемонстрируйте связь метода наименьших квадратов с методом максимального правдоподобия.
3. Что такое переобучение?
4. В чем заключается задача классификации?
5. Что такое нейронная сеть?
6. Что такое деревья решений?

Критерии оценок выполнения заданий/задач (каждая задача оценивается в 2 балла)

Решена полностью	2
Решена основная часть задачи, или задача решена с недочетами	1,5
Решена задача наполовину	1
Сделан первый этап в решении задачи	0,5
Нет решения	0

Суммарная оценка выполнения заданий/задач

Количество баллов	Оценка
4	Отлично
3,5	Очень хорошо
3	Хорошо
2-2,5	Удовлетворительно
0,5-1,5	Неудовлетворительно
0	Плохо

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

- 1) Воронцов К.В. Машинное обучение. Курс лекций.
<http://www.intuit.ru/studies/courses/13844/1241/info>

б) Дополнительная литература

- 1) Золотых Н.Ю. Машинное обучение. Курс лекций. Нижний Новгород: ННГУ, 2007.
<http://www.uic.nnov.ru/~zny/ml>

в) программное обеспечение и Интернет-ресурсы

- 1) Международный открытый ресурс и программное обеспечение для машинного обучения

- Scikit-Learn: Machine Learning in Python: www.scikit-learn.org
2) The R Project for Statistical Computing <https://www.r-project>
3) Welcome to Python.org <https://www.python.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 Алгебры, геометрии и дискретной математики.

Программа составлена в соответствии с требованиями ОС ВО ННГУ с учетом рекомендаций ФГОС ВО по направлению 02.04.02. – Фундаментальная информатика и информационные технологии.

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

Рецензенты: _____

Заведующий кафедрой _____ Золотых Н.Ю.

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики от 30.11.2022 года, протокол № 3.