

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 №13 от 30.11.2022 г.

Рабочая программа дисциплины
Машинное обучение и анализ данных

Уровень высшего образования
бакалавриат

Направление подготовки
**020302 Фундаментальная информатика и информационные
технологии**

Направленность образовательной программы
Инженерия программного обеспечения

Форма обучения
очная

Нижегород
2023

1. Место и цели дисциплины в структуре ООП

Место дисциплины в структуре ООП

Данная дисциплина читается в первом семестре магистратуры (Б1.В.03 – часть, формируемая участниками образовательных отношений).

Целями освоения дисциплины являются 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.

В результате освоения дисциплины студенты должны

Знать:

- 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 обратного распространения ошибки для обучения нейронной сети, машина опорных векторов, деревья решений, градиентный бустинг, баггинг, случайные деревья);
- алгоритмы кластеризации (методы из теории графов, методы центров тяжести и медиан, ЕМ-метод). Подходы к решению задач иерархической кластеризации;
- экспериментальные методы оценки качества обучения;
- основы теории Вапника–Червоненкиса.

Уметь:

- использовать имеющиеся знания для решения практических задач машинного обучения;
- оценивать на практике качество обучения используемых моделей.

Иметь навыки

- решения практических задач с использованием методов машинного обучения;
- использования среды статистических вычислений R или библиотеки ScikitLearn для решения задач 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	

ПК-2 Способен к применению общенаучных базовых знаний математических и естественных наук, фундаментальной информатики и информационных технологий; способность применять в профессиональной деятельности современные языки программирования и методы параллельной обработки данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии.	ПК-2.1. Знает основы научно-исследовательской деятельности в области информационных технологий, имеет научные знания в теории информационных систем.	ЗНАТЬ основные концептуальные и теоретические модели машинного обучения, основные алгоритмы и методы машинного обучения основы языка Python или среды вычислений R	Собеседование
	ПК-2.2. Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности	УМЕТЬ использовать методы машинного обучения на практике, оценивать качество методов работать с библиотекой Scikit-Learn или средой для статистических вычислений R	Контрольная работа (текущий контроль)
	ПК-2.3. Имеет практический опыт научно-исследовательской деятельности в области информационных технологий.	ИМЕЕТ практический опыт разработки и анализа концептуальных и теоретических моделей машинного обучения	Контрольная работа (текущий контроль)

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

	Очная форма обучения
Общая трудоемкость	3 ЗЕТ
Часов по учебному плану	108
в том числе	
аудиторные занятия (контактная работа):	33
- занятия лекционного типа	32
- занятия семинарского типа	-
- занятия лабораторного типа	-
- текущий контроль (КСР)	1
самостоятельная работа	75

Промежуточная аттестация – зачет	
----------------------------------	--

3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины	Всего (часы)	В том числе				Самостоятельная работа обучающегося, часы
		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	6	2			2	4
2. Вероятностная постановка задачи обучения по прецедентам	7	2			2	5
3. Экспериментальные методы оценки качества обучения	7	2			2	5
4. Метод наименьших квадратов для решения задачи восстановления регрессии	7	2			2	5
5. Проблема переобучения при решении задачи восстановления регрессии	7	2			2	5
6. Метод ближайших соседей для решения задачи классификации	7	2			2	5
7. Наивный байесовский классификатор	7	2			2	5
8. Линейный дискриминантный анализ	7	2			2	5
9. Нейронные сети. Персептрон Розенблатта	7	2			2	5
10. Машина опорных векторов	7	2			2	5
11. Деревья решений	7	2			2	5
12. Ансамбли решающих правил	7	2			2	5
13. Обучение без учителя	7	2			2	5
14. Иерархическая кластеризация	7	2			2	5
15. Основы теории Вапника–Червоненкиса	10	4			4	6
Текущий контроль (КСР)	1				1	
Промежуточная аттестация – зачет						
Итого	108	32		0	33	75

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

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 лекционных занятиях. Экспресс – опрос оценивается оценками «Зачтено» – «Не зачтено».

4.2. 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.

4.3. Подготовка к промежуточной аттестации в форме экзамена

В качестве методических материалов при подготовке к зачету 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	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.

	я от ответа						
<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 ПК-2

1. Какая из следующих задач является задачей обучения без учителя?
 - a. восстановление регрессии
 - b. классификация
 - c. кластеризация (+)
2. Какая из следующих задач является задачей обучения с учителем?
 - a. классификация (+)
 - b. кластеризация
 - c. восстановление функции распределения
3. Переобучением называется
 - a. явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много больше ошибки на обучающей выборке (+)
 - b. явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много меньше ошибки на обучающей выборке
 - c. 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 ПК-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 $P(Y = 1 \mid x_1 = 1, x_1 = 2)$
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. Контрольные вопросы

Вопрос	Код компетенции (согласно РПД)
1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор.	ПК-2
2. Экспериментальные методы оценки качества обучения.	ПК-2
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-2
4. Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.	ПК-2
5. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.	ПК-2
6. Наивный байесовский классификатор.	ПК-2
7. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.	ПК-2
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	ПК-2

борьбы с переобучением. Понятие о глубоких нейронных сетях.	
9. Машина опорных векторов. Ядра и спрямляющие пространства.	ПК-2
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-2
11. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосования. 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-2
12. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. EM-алгоритм.	ПК-2
13. Иерархическая кластеризация. Агломеративные и разделяющие методы.	ПК-2
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-2
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-2

6. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература:

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

б) дополнительная литература:

1. 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 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 02.03.02 Фундаментальная информатика и информационные технологии.

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

Рецензент: _____

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

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики 13.12.2023 года, протокол № 3