

Министерство науки и высшего образования Российской Федерации  
федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Кубанский государственный университет»  
факультет математики и компьютерных наук

УТВЕРЖДАЮ  
Проректор по учебной работе,  
качеству образования – первый  
проректор  
\_\_\_\_\_ Кагуров А.



«29» мая 2020 г.

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Б1.В.ДВ.04.01 Математические методы машинного обучения**

Направление подготовки: 01.05.01 Фундаментальные математика и механика

Специализация: Вычислительная механика и компьютерный инжиниринг

Форма обучения: очная

Квалификация: Математик. Механик. Преподаватель

Краснодар 2020

Рабочая программа дисциплины Б1.В.ДВ.8 МАТЕМАТИЧЕСКИЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 01.05.01 Фундаментальные математика и механика

Программу составил:

Голуб М.В., доцент, доктор физ.-мат. наук \_\_\_\_\_

Рабочая программа дисциплины Б1.В.ДВ.8 МАТЕМАТИЧЕСКИЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ утверждена на заседании кафедры теории функций  
протокол № 8 «17» марта 2020 г.

Заведующий кафедрой теории функций Голуб М.В. \_\_\_\_\_

Рабочая программа обсуждена на заседании кафедры теории функций  
протокол № 8 «17» марта 2020 г.

Заведующий кафедрой теории функций Голуб М.В. \_\_\_\_\_

Утверждена на заседании учебно-методической комиссии факультета математики и компьютерных наук  
протокол № 2 «30» апреля 2020 г.

Председатель УМК факультета Шмалько С.П. \_\_\_\_\_

Рецензенты:

Гусаков Валерий Александрович, канд. физ. – мат. наук,  
директор ООО «Просвещение – Юг»

Засядко Ольга Владимировна, доцент кафедры информационных образовательных технологий, канд. физ. - мат. наук, доцент

## **1 Цели и задачи изучения дисциплины.**

### **1.1 Цель освоения дисциплины.**

Цель дисциплины «Математические методы машинного обучения» – обучение студентов базовым понятиям машинного обучения, формированию навыков разработки основных моделирующих алгоритмов машинного обучения и реализации их на базе языков и пакетов прикладных программ моделирования; овладение современным методам машинного обучения для дальнейшего использования в других областях математического знания и дисциплинах естественнонаучного содержания.

### **1.2 Задачи дисциплины.**

- изучить математическую постановку задач обучения по прецедентам;
- сформировать навыки разработки моделирующих алгоритмов при решении комплекса задач анализа данных;
- получить практические навыки реализации обучающих алгоритмов на базе языков и пакетов прикладных программ моделирования.

### **1.3 Место дисциплины (модуля) в структуре образовательной программы.**

Дисциплина «Математические методы машинного обучения» относится к вариативной части профессионального цикла Блока1 "Дисциплины (модули)" учебного плана (Б1.В.ДВ). Для успешного освоения дисциплины обучающийся должен владеть знаниями, умениями и навыками по программе дисциплин «Теория вероятностей», «Математическая статистика», «Математическое моделирование», «Технология программирования и работа на электронно-вычислительной машине (ЭВМ)».

### **1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы.**

Изучение данной учебной дисциплины направлено на формирование у обучающихся общепрофессиональных, профессиональных компетенций: ПК-3, ОПК-1.

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
1.	ОПК-4	способностью находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем	общие характеристики современных языков и пакетов прикладных программ моделирования	реализовывать алгоритмы на базе языков и пакетов прикладных программ моделирования	навыками разработки моделирующих алгоритмов и реализации их на базе языков и пакетов прикладных программ моделирования;

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
2	ПК-3	способностью создавать и исследовать новые математические модели явлений реального мира, сред, тел и конструкций	классы задач обучения по прецедентам; принципы построения моделирующ их алгоритмов.	применять на практике алгоритмы машинного обучения; обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи	базовым инструментарием машинного обучения; методами анализа алгоритмов, методами сведения задач к стандартным задачам

## 2. Структура и содержание дисциплины.

### 2.1 Распределение трудоёмкости дисциплины по видам работ с формой контроля в 7 семестре – зачет.

Общая трудоёмкость дисциплины составляет 3 зачетных единиц (108 часов, из них – 60,2 часа контактной работы (в том числе: лекционных 28 ч., лабораторных 28 ч.; 4 часов КСР, 0,2 ч ИКР), 47,8 часа самостоятельной работы).

Вид учебной работы	Всего часов	Семестры (часы)
		7
<b>Контактная работа, в том числе:</b>	<b>60,2</b>	<b>60,2</b>
<b>Аудиторные занятия (всего):</b>	<b>56</b>	<b>56</b>
Занятия лекционного типа	28	28
Лабораторные занятия	28	28
Занятия семинарского типа (семинары, практические занятия)	-	-
<b>Иная контактная работа:</b>	<b>4,2</b>	<b>4,2</b>
Контроль самостоятельной работы (КСР)	4	4
Промежуточная аттестация (ИКР)	0,2	0,2
<b>Самостоятельная работа, в том числе:</b>	<b>47,8</b>	<b>47,8</b>
Выполнение индивидуальных заданий (подготовка сообщений, презентаций)	-	-
Проработка учебного материала	47,8	47,8
<b>Контроль:</b>	<b>-</b>	<b>-</b>
Подготовка к текущему контролю	-	-
<b>Общая трудоёмкость</b>	<b>час.</b>	<b>108</b>
	<b>в том числе контактная работа</b>	<b>60,2</b>
	<b>зач. ед</b>	<b>3</b>

## 2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам (темам) дисциплины.

Разделы дисциплины, изучаемые в 7 семестре

№	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение в машинное обучение	6	2		2	2
2.	Байесовские методы классификации	14,4	4		4	6,4
3.	Метрические методы классификации	14	4		4	6
4.	Линейные и нелинейные методы классификации	20	6		6	8
5.	Методы восстановления регрессии	20	6		6	8
6.	Искусственные нейронные сети	17,8	4		4	9,4
7.	Кластеризация	12	2		2	8
	<i>Итого по дисциплине:</i>		28		28	47,8

## 2.3 Содержание разделов (тем) дисциплины:

### 2.3.1 Занятия лекционного типа.

№ раздела	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1	Введение в машинное обучение.	Объекты и факторы. Модель алгоритмов и метод обучения. Функционал качества. Проблема переобучения и понятие обучающей способности. Примеры прикладных задач.	Устные опросы.
2	Байесовские методы классификации.	Вероятностная постановка задачи классификации. Оптимальное байесовское решающее правило. Наивный байесовский классификатор. Линейный и квадратичный дискриминантный анализ. EM-алгоритм. Смеси многомерных нормальных распределений.	Устные опросы.
3	Метрические методы классификации	Метод рассуждения по прецедентам. Обобщенный метрический классификатор. Метод ближайшего соседа. Алгоритм ближайших $k$ соседей. Метод парзеновского окна. Алгоритм STOLP для выбора эталонных объектов.	Устные опросы.

4	Линейные методы классификации	Линейная модель классификации. Метод стохастического градиента. Логистическая регрессия. Метод опорных векторов. Метод релевантных векторов.	Устные опросы.
5	Методы восстановления регрессии	Метод наименьших квадратов. Линейная регрессия. Проблема мультиколлинеарности. Гребневая регрессия. Метод главных компонент. Нелинейные методы восстановления регрессии. Итерационный взвешенный МНК. Методы опорных и релевантных векторов в задачах регрессии.	Устные опросы.
6	Искусственные нейронные сети	Вычислительные возможности нейронных систем. Многослойные нейронные сети. Метод обратного распространения ошибок. Выбор начального приближения. Оптимизация структуры сети.	Устные опросы.
7	Кластеризация	Функционалы качества кластеризации. Статистические алгоритмы кластеризации. Иерархическая кластеризация.	Устные опросы.

### 2.3.2 Занятия семинарского типа.

Не предусмотрены.

### 2.3.3 Лабораторные занятия. 7 семестр

№	Наименование раздела	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Введение в машинное обучение.	Реализация моделей Data Mining в среде R. Основные шаги построения и верификации моделей. Многомерный статистический анализ. Проецирование многомерных данных на плоскости.	Обсуждение домашних заданий.
2.	Байесовские методы классификации.	Построение наивного байесовского классификатора. Классификация в линейном дискриминантном пространстве. Построение классифицирующих функций.	Обсуждение домашних заданий. Контрольная работа 1.

3.	Метрические методы классификации	Метрики при решении задач классификации. Построение обобщенного метрического классификатора. Ирисы Фишера и метод k ближайших соседей. Алгоритм DBSCAN.	Обсуждение домашних заданий. Защита результатов выполнения индивидуальных заданий
4.	Линейные методы классификации	Бинарные классификаторы и логистическая регрессия. Модель мультиномиального логита. Ядерные функции машины опорных векторов. Метод опорных векторов. Сравнение метода релевантных векторов по сравнению с опорными. Нелинейные классификаторы в R.	Обсуждение домашнего задания. Контрольная работа 2
	Методы восстановления регрессии	Выбор оптимального набора предикторов. Пошаговые процедуры построения регрессии. Генетический алгоритм. Реализация метода главных компонент. Оценивание нелинейных по параметрам регрессий нелинейным МНК	Обсуждение домашних заданий. Защита результатов выполнения индивидуальных заданий
	Искусственные нейронные сети	Классификаторы на основе искусственных нейронных сетей. Двуслойный персептрон Розенблата и логистическая регрессия. Настройка параметров нейронной сети.	Обсуждение домашних заданий. Защита результатов выполнения индивидуальных заданий
	Кластеризация	Реализация алгоритмов кластеризации, основанных на разделении. Оценка качества кластеризации. Иерархическая кластеризация.	Обсуждение домашнего задания. Контрольная работа 3

**2.3.4 Примерная тематика курсовых работ (проектов) – курсовая работа не предусмотрена**

**2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)**

**7 семестр**

№	Наименование раздела	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3

1.	Концепция метода конечных элементов, пространства функций, формулировки задач.	[1]- [5] (из списка 5.1), или [1]-[4] (из списка 5.2)
2.	Построение математических моделей механики сплошных сред.	[1]- [5] (из списка 5.1), или [1]-[4] (из списка 5.2)
3.	Исследование математических моделей при решении прикладных задач	[1]- [5] (из списка 5.1), или [1]-[4] (из списка 5.2)
4.	Вычислительный эксперимент и его роль	[1]- [5] (из списка 5.1), или [1]-[4] (из списка 5.2)

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля. Для лиц с нарушениями слуха:
- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

### **3. Образовательные технологии.**

При изучении данного курса не используются.

### **4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации.**

#### **4.1 Фонд оценочных средств для проведения текущего контроля.**

##### **Примерный перечень заданий и контрольных вопросов**

1. Дайте определение эффективности оценки параметра.
2. Какие из нижеперечисленных факторов приводят к смещенности МНК-оценок:
  - гетероскедастичность;
  - мультиколлинеарность;
  - пропущенные существенные переменные;
  - включенные несущественные переменные.
3. Перечислите условия Гаусса-Маркова для линейной парной регрессии.

4. Если коэффициент уравнения регрессии  $\beta$  статистически значим, то
  - a)  $\beta \neq 0$ ;
  - b)  $\beta > 0$ ;
  - c)  $0 < \beta < 1$ ;
  - d)  $\beta > 0$ ;
  - e)  $|\beta| > 1$ .
5. Для модели парной регрессии  $y_t = \alpha + \beta x_t + \varepsilon_t$  обозначим через  $a, b$  МНК-оценки параметров  $\alpha, \beta$ . Пусть  $c$  – положительная константа. Обозначим  $y_{1t} = c y_t$ . Пусть  $a_1, b_1$  – МНК-оценки параметров  $\alpha_1, \beta_1$  в модели  $y_{1t} = \alpha_1 + \beta_1 x_t + \varepsilon_{1t}$ . Выразите  $a_1, b_1$  через  $a, b$  и  $c$ .
6. Дана оценка регрессии  $\ln \hat{y}_t = 1.45 + 0.27x_{1t} + 1.68 \ln x_{2t}$ , где  $y$  – цена за квартиру,  $x_{1t}$  – лифт,  $x_{2t}$  – площадь квартиры. Дайте интерпретацию коэффициентов.
7. Оцененная зависимость расходов на жилищное строительство в США в 1977–2000 годах (в млрд. долл.) от времени  $t = 1$  в 1977 г.,  $t = 2$  в 1978 г. и т.д. с учетом сезонных факторов ( $d_i = 1$ , если наблюдение относится к  $i$ -му кварталу, иначе 0), имеет вид:  $y_t = 14 + 3d_2 + 4d_3 + 3d_4 - 0.5t$ . Оцените жилищные расходы в первом и втором квартале 1980 г.

#### 4.2 Фонд оценочных средств для проведения промежуточной аттестации.

##### *Вопросы к зачету*

1. Общая формула байесовского классификатора.
2. Три подхода к восстановлению плотности распределения по выборке.
3. Что такое наивный байесовский классификатор?
4. Многомерное нормальное распределение. Вывести формулу квадратичного дискриминанта. При каком условии он становится линейным?
5. На каких предположениях основан линейный дискриминант Фишера?
6. Последовательное добавление компонент в EM-алгоритме, основная идея алгоритма.
7. Приведите пример метрического алгоритма классификации, который одновременно является байесовским классификатором.
8. Приведите пример метрического алгоритма классификации, который одновременно является линейным классификатором.
9. Зачем нужен отбор опорных объектов в метрических алгоритмах классификации?
10. Обоснование логистической регрессии.
11. Две мотивации и постановка задачи метода опорных векторов.
12. Какое ядро порождает полиномиальные разделяющие поверхности?
13. Постановка задачи многомерной линейной регрессии. Матричная запись.
14. «Проблема мультиколлинеарности» в задачах многомерной линейной регрессии.
15. Сравнить гребневую регрессию и лассо.
16. Сведение задачи многомерной нелинейной регрессии к последовательности линейных задач.
17. Метод обратного распространения ошибок. Основная идея.
18. Как выбирать число слоёв в градиентных методах настройки нейронных сетей?
19. В чём заключается метод оптимального прореживания нейронной сети?
20. Основные типы кластерных структур.
21. Какие существуют функционалы качества кластеризации и для чего они применяются?
22. Основные отличия алгоритма k-средних и EM-алгоритма.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

- при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на зачете;
- при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;
- при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа. Для лиц с нарушениями слуха:
- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

## **5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля).**

### **5.1 Основная литература:**

1. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования : учебное пособие / В. В. Вьюгин. — Москва: МЦНМО, 2014. — 304 с. — ISBN 978-5-4439-2014-6. — Режим доступа: <https://e.lanbook.com/book/56397>
2. Шалев-Шварц, Ш. Идеи машинного обучения: учебное пособие / Ш. Шалев-Шварц, Бен-Давид Ш. ; перевод с английского А. А. Слинкина. — Москва: ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060-673-5. — Режим доступа: <https://e.lanbook.com/book/131686>
3. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва: ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7. — Режим доступа: <https://e.lanbook.com/book/69955>

### **5.2 Дополнительная литература:**

1. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. – М.: Финансы и статистика, 1989.
2. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. – М.: Наука, 1979.
3. Bishop C. M. Pattern Recognition and Machine Learning. – Springer, Series: Information Science and Statistics, 2006. – 740 pp.
4. K.P. Murphy. Machine learning: a probabilistic perspective. Adaptive computation and machine learning series/ Murphy K.P. – Massachusetts Institute of Technology, 2012.

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах «Лань» и «Юрайт».

## **6. Перечень ресурсов информационно телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля).**

[github.com](https://github.com)

## **7. Методические указания для обучающихся по освоению дисциплины.**

Самостоятельная работа студента является необходимой и крайне важной при изучении любого теоретического или практического учебного курса и должна быть правильно организована. Прежде всего, необходимо, чтобы эта работа была систематической и регулярной. Самостоятельная работа делится между теоретической частью курса и практической, но это деление не носит формального характера, поскольку решение практических задач предусматривает знание основных теоретических понятий и методов, а теоретические знания в свою очередь не могут усваиваться без практической работы с теоретическими конструкциями.

При подготовке к практическому занятию студенту целесообразно познакомиться сначала с теоретическими понятиями, относящимися к данному разделу, чтобы уяснить для себя смысловую часть работы. Для этого рекомендуется прочитать лекции или учебники, в которых освещаются соответствующие вопросы. Естественно, студенту необязательно использовать лишь литературу, указанную в библиографии, но на начальных стадиях изучения материала это делать желательно. Со временем расширение использования литературных источников можно лишь приветствовать. Перед решением домашних задач студенту целесообразно познакомиться сначала с содержанием предыдущего занятия, уяснить для себя методы решения задач рассматриваемого типа. При этом у студента естественно возникают затруднения и вопросы, которые он может задать преподавателю на следующем практическом занятии. Любое практическое занятие начинается с разборов вопросов и затруднений по домашнему заданию. Форма практических занятий, особенно занятий лабораторных, предусматривает диалог между студентами и преподавателем. Практика показывает, что студенты охотно прибегают к прямому диалогу с преподавателем и умеют извлечь для себя пользу из соответствующего диалога. Каждая большая тема заканчивается итоговой контрольной работой с выставлением оценки. Студент должен получить по каждой контрольной работе хотя бы удовлетворительную оценку, иначе он получает дополнительное задание с обязательным условием отработки неудовлетворительной оценки по соответствующей контрольной работе. Эти отработки принимаются преподавателем, ведущим практические занятия в течение всего семестра.

На зачете студенту предлагается билет с двумя теоретическими вопросами. По получении билета студент имеет возможность в течение 15 минут почитать конспект своих лекций, после чего в течение тридцати минут он должен письменно изложить теоретический материал по билету. Практика показывает, что студент, не изучавший материал, не может действительно воспользоваться лекциями при подготовке к письменному ответу. Наоборот, даже сильному студенту трудно изложить теоретический материал без предварительного просмотра материала в течение 15 минут перед письменным ответом.

## **8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.**

### **8.1 Перечень информационных технологий.**

1. Мультимедиа и коммуникационные технологии.
2. Элементы дистанционных технологий.
3. Мировые информационные образовательные ресурсы.
4. Аудиовизуальные и интерактивные средства обучения.
5. Мобильное обучение.

6. Облачные технологии.

### 8.2 Перечень необходимого программного обеспечения.

1. Microsoft Windows 8, 10 "№73–АЭФ/223-ФЗ/2018 Соглашение Microsoft ESS 72569510"
2. Microsoft Office Professional Plus "№73–АЭФ/223-ФЗ/2018 Соглашение Microsoft ESS 72569510"
3. Acrobat DC №79-АЭФ/223-ФЗ/2017
4. RStudio Version 1.1.456 – © 2009-2018 RStudio, Inc
5. Свободная программная среда вычислений R.

### 8.3 Перечень информационных справочных систем:

1. Электронная библиотечная система издательства «Лань» – тематические коллекции (<http://e.lanbook.com>)
2. Электронная библиотечная система «Университетская библиотека онлайн» – базовая коллекция ([www.biblioclub.ru](http://www.biblioclub.ru))

## 9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

№	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащённость
1.	Лекционные занятия	Лекционная аудитория, специально оборудованная мультимедийными демонстрационными комплексами, учебной мебелью
2.	Лабораторные занятия	Помещение для проведения лабораторных занятий оснащённое учебной мебелью, доской маркером или мелом
3.	Групповые (индивидуальные) консультации	Помещение для проведения групповых (индивидуальных) консультаций, учебной мебелью, доской маркером или мелом
4.	Текущий контроль, промежуточная аттестация	Помещение для проведения текущей и промежуточной аттестации, оснащённое учебной мебелью.
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащённый компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета