

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям



О.Г. Прохоренко

«05» июля 2023 г.

Регистрационный № УД – 12964/уч.

Математические основы машинного обучения

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-31 03 01 Математика (по направлениям)

Направления специальности:

1-31 03 01-01 Математика (научно-производственная деятельность)

2023 г.

Учебная программа составлена на основе ОСВО 1-31 03 01-2021, типового учебного плана, регистрационный № G31-1-011/пр.-тип. от 31.03.2021, учебных планов: № G 31-1-003/уч. от 25.05.2021, № G 31-1-061/уч. ин. от 31.05.2021

СОСТАВИТЕЛИ:

С.А. Бондарев, доцент кафедры теории функций Белорусского государственного университета, кандидат физико-математических наук, доцент;

Н.В. Бровка, заведующий кафедрой теории функций Белорусского государственного университета, доктор педагогических наук, профессор;

В.Г. Кротов, профессор кафедры теории функций Белорусского государственного университета, доктор физико-математических наук, профессор;

Т.С. Мардвилко, доцент кафедры теории функций Белорусского государственного университета, кандидат физико-математических наук, доцент;

Н.Б. Яблонская, доцент кафедры общей математики и информатики Белорусского государственного университета, кандидат физико-математических наук, доцент.

РЕЦЕНЗЕНТЫ:

В.В. Гороховик, гл. научный сотрудник отдела нелинейного и стохастического анализа Института математики Национальной Академии Наук Республики Беларусь, доктор физико-математических наук, профессор, член-корреспондент Национальной Академии наук Беларуси.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории функций БГУ
(протокол № 18 от 07.06.2023);

Научно-методическим советом БГУ
(протокол № 9 от 29.06.2023)

Заведующий
кафедрой теории функций



Н.В. Бровка

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины «Математические основы машинного обучения» – создание базы для освоения основных понятий и методов современной математики.

Образовательная цель: изложение основ машинного обучения.

Развивающая цель: формирование у студентов основ математического мышления, знакомство с методами машинного обучения, изучение основных алгоритмов машинного обучения и реализация их на практике.

Задачи учебной дисциплины «Математические основы машинного обучения»:

1. Формирование у студентов понятия обучения по прецедентам;
2. Изучение классических алгоритмов машинного обучения;
3. Изучение основ глубокого обучения (нейронных сетей);
4. Использование алгоритмов машинного обучения для решения конкретных прикладных задач.

Место учебной дисциплины

Учебная дисциплина «Математические основы машинного обучения» относится к дисциплинам специализации компонента учреждения высшего образования.

Связи с другими учебными дисциплинами

Учебная дисциплина «Математические основы машинного обучения» тесно связана с такими дисциплинами как «Математический анализ», «Экстремальные задачи», «Численные методы», «Теория вероятностей и математическая статистика».

Требования к компетенциям

Освоение учебной дисциплины «Математические основы машинного обучения» должно обеспечить формирование следующих универсальных и базовых профессиональных компетенций:

универсальные компетенции:

УК-1. Владеть основами исследовательской деятельности, осуществлять поиск, анализ и синтез информации.

базовые профессиональные компетенции:

БПК-2. Использовать понятия и методы вещественного, комплексного и функционального анализа и применять их для изучения моделей окружающего мира.

В результате освоения учебной дисциплины студент должен:

знать:

- основные понятия и результаты теории статистического обучения машин;
- методы доказательств и алгоритмы решения задач машинного обучения;
- новейшие достижения в области машинного обучения и их приложения в прикладных задачах;

уметь:

- использовать основные результаты машинного обучения в практической деятельности;
- использовать теоретические и практические навыки применения теории статистического обучения по прецедентам в математике;

владеть:

- основными классическими алгоритмами машинного обучения;
- навыками самообразования и способами использования аппарата машинного обучения для проведения математических и междисциплинарных исследований.

Структура учебной дисциплины

Учебная дисциплина «Математические основы машинного обучения» изучается в 6 семестре очной формы получения высшего образования. Всего на ее изучение отведено:

- в очной форме получения высшего образования: 120 часов, в том числе 68 аудиторных часа, из них: лекции – 62 часов, управляемая самостоятельная работа – 6 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Основные понятия теории статистического обучения по прецедентам

Постановка основных задач машинного обучения. Объекты и признаки. Типы задач: классификация, регрессия, ранжирование. Методы обучения: с учителем, без учителя, с частичным привлечением учителя, обучение с подкреплением. Основные понятия: модель алгоритмов, функционал качества, минимизация эмпирического риска, обобщающая способность, переобучение. Отличие от интерполяции, пример Рунге. Примеры прикладных задач.

Тема 2. Линейные модели

Линейная регрессия. Метод наименьших квадратов.

Линейная классификация, аппроксимация пороговой функции потерь.

Метод градиентного спуска, стохастического градиентного спуска. Проблема мультиколлинеарности и переобучения.

Тема 3. Вероятностная постановка задач машинного обучения

Вероятностная постановка задач классификации. Принцип максимума правдоподобия. Регуляризация, ее вероятностная интерпретация – гауссовский и лапласовский регуляризаторы. Логистическая регрессия.

Тема 4. Нейронные сети

Модель нейрона МакКаллока-Питтса. Функции активации. Нейронные сети как универсальный аппроксиматор. Обучение нейронных сетей – метод обратного распространения ошибки. Проблема взрыва и затухания градиентов. Методы Dropout и BatchNorm.

Тема 5. Метрические методы

Гипотезы компактности и непрерывности. Обобщенный метрический классификатор. Метод ближайших соседей kNN и его обобщения. Метод окна Парзена. Метод потенциальных функций, связь с линейными моделями.

Непараметрическая регрессия. Формула Надарая-Ватсона.

Тема 6. Метод опорных векторов

Оптимальная разделяющая гиперплоскость, понятие зазора между классами. Случай линейной разделимости. Сведение к задаче квадратичного программирования. Понятие опорных векторов. Функции ядра, спрямляющее пространство, теорема Мерсера. Способы конструктивного построения ядер, примеры.

Тема 7. Многомерная линейная регрессия

Многомерная линейная регрессия, метод наименьших квадратов, геометрический смысл.

Сингулярное разложение, мультиколлинеарность и переобучение. Гребневая регрессия.

Метод главных компонент. Понижение размерности, задачи низкоранговых матричных разложений.

Тема 8. Нелинейная регрессия

Метод Ньютона-Рафсона, метод Ньютона-Гаусса. Логистическая регрессия, метод IRLS.

Обобщенная линейная модель. Экспоненциальное семейство распределений.

Тема 9. Критерии выбора моделей и отбора признаков

Критерии качества классификации: ROC-кривая, точность, полнота, PR-кривая.

Внутренние и внешние критерии. Кросс-проверка.

Информационный критерий Акаике, байесовский информационный критерий, оценка Вапника-Червоненкиса, VC-размерность.

Задача отбора признаков. Полный перебор, жадные алгоритмы, поиск в глубину и ширину.

Тема 10. Логические методы

Понятие логической закономерности. Параметрические семейства закономерностей: конъюнкции пороговых правил. Переборные алгоритмы синтеза конъюнкций.

Статистический критерий информативности. Тест Фишера.

Решающее дерево, жадная стратегия. Алгоритм ID3. Случайный лес. Деревья регрессии.

Тема 11. Методы ансамблирования

Базовый алгоритм, простое голосование.

Бэггинг, случайный лес, AdaBoost.

Бустинг, градиентный бустинг, стохастический градиентный бустинг. Варианты бустинга. Стэкинг.

Тема 12. Восстановление плотности распределения

Параметрическое оценивание плотности, многомерное нормальное распределение. Выборочные оценки параметров. Непараметрическая оценка плотности, смесь распределений, EM-алгоритм.

Тема 13. Байесовская теория классификации

Оптимальный байесовский классификатор. Генеративные и дискриминативные модели. Наивный байесовский классификатор. Линейный наивный байесовский классификатор в случае экспоненциального семейства распределений.

Нормальный дискриминатный анализ. Линейный дискриминант Фишера.

Смесь многомерных нормальных распределений. Сеть радиальных базисных функций.

Тема 14. Кластеризация

Постановка задачи. Примеры. Алгоритмы k-means, DBSCAN.

Построение дендрограммы. Определение числа кластеров. Другие алгоритмы кластеризации.

Тема 15. Глубокие нейронные сети

Полносвязные нейронные сети. Сверточные сети для изображений, ResNet.

Рекуррентные нейронные сети.

Перенос обучения (transfer learning) и дистилляция нейронных сетей.

Модели внимания и трансформеры.

Тема 16. Автоматическая обработка текстов

Векторные представления текста. Задача машинного перевода. Тематическое моделирование. Латентное размещение Дирихле.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная форма получения высшего образования с применением дистанционных образовательных технологий (ДОТ)

Номер раздела, те- мы	Название раздела, темы	Количество аудиторных часов						Количество часов УСР (аудиторный контроль)	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное			
1	2	3	4	5	6	7	8	9	
1	Основные понятия теории ста- тистического обучения по пре- цедентам	2							
2	Линейные модели	4					1	Отчет по лабора- торной работе	
3	Вероятностная постановка за- дач машинного обучения	4							
4	Нейронные сети	4					1	Отчет по лабора- торной работе	
5	Метрические методы	4					1	Отчет по лабора- торной работе	
6	Метод опорных векторов	4							
7	Многомерная линейная регрес- сия	4							
8	Нелинейная регрессия	4					1	Отчет по лабора- торной работе	
9	Критерии выбора моделей и от- бора признаков	4							
10	Логические методы	4							
11	Методы ансамблирования	4					1	Отчет по лабора-	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

- 1 Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014. — 739 p.
- 2 Bishop C. M. Pattern Recognition and Machine Learning. — Springer, 2006. — 738 p.
- 3 Мерков А.Б. Распознавание образов. Введение в методы статистического обучения. 2011. 256 с.
- 4 Мерков А.Б. Распознавание образов. Построение и обучение вероятностных моделей. 2014. 238 с.
- 5 Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.
- 6 Я.Лекун. Как учится машина. Революция в области нейронных сетей и глубокого обучения. (Библиотека Сбера: Искусственный интеллект). — М.: Альпина нон-фикшн, 2021
- 7 Таулли, Т. Основы искусственного интеллекта. Нетехническое введение = Artificial Intelligence Basics. A Non-Technical Introduction / Том Таулли ; [пер. с англ. Андрея Логунова]. - Санкт-Петербург : БХВ-Петербург, 2021. - 288 с. - URL: <https://ibooks.ru/bookshelf/385769>.
- 8 Дайзенрот, М. П. Математика в машинном обучении: докопайся до сути / М. П. Дайзенрот, А. А. Фейзал, Чен Сунь Он; [пер. с англ. С. Череников]. - Санкт-Петербург: Минск: Питер, 2024. - 507 с.
- 9 Воронов, М. В. Системы искусственного интеллекта : учебник и практикум для студентов высших учебных заведений, обучающихся по ИТ и математическим направлениям / М. В. Воронов, В. И. Пименов, И. А. Небаев. - Москва : Юрайт, 2022. - 256 с.

Перечень дополнительной литературы

- 10 Червоненкис А. Я. Компьютерный анализ данных. Лекции Школы анализа данных Яндекс, 2009. 260 стр..
- 11 Шалев-Шварц Ш., Бен-Давид Ш. Идеи машинного обучения. ДМК-Пресс, 2019. 436 с.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Объектом диагностики компетенций студентов являются знания, умения, полученные ими в результате изучения учебной дисциплины. Выявление учебных достижений студентов осуществляется с помощью мероприятий текущего контроля и промежуточной аттестации.

Для диагностики компетенций могут использоваться следующие средства текущего контроля: отчет по лабораторной работе.

Формой промежуточной аттестации по дисциплине учебным планом предусмотрен – экзамен.

При формировании итоговой отметки используется рейтинговая система оценки знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая система предусматривает использование весовых коэффициентов в ходе проведения контрольных мероприятий текущей аттестации.

Примерные весовые коэффициенты, определяющие вклад текущей аттестации в отметку при прохождении промежуточной аттестации:

Формирование отметки за текущую аттестацию:

– Отчет по лабораторной работе (по всем лабораторным работам) – 100 %.

Итоговая отметка по дисциплине рассчитывается на основе отметки текущей аттестации (рейтинговой системы оценки знаний) - 40% и экзаменационной отметки - 60%.

Примерный перечень заданий для управляемой самостоятельной работы студентов

Тема 2. Линейные модели (1 ч).

Примерный перечень заданий

Линейная регрессия. Метод наименьших квадратов.

Линейная классификация, аппроксимация пороговой функции потерь.

Метод градиентного спуска, стохастического градиентного спуска. Проблема мультиколлинеарности и переобучения.

Форма контроля – отчет по лабораторной работе.

Тема 4. Нейронные сети (1 ч).

Примерный перечень заданий

Обучение нейронных сетей – метод обратного распространения ошибки.

Проблема взрыва и затухания градиентов. Методы Dropout и BatchNorm.

Форма контроля – отчет по лабораторной работе.

Тема 5. Метрические методы (1 ч).

Примерный перечень заданий

Метод ближайших соседей kNN и его обобщения.

Непараметрическая регрессия. Формула Надарая-Ватсона.

Форма контроля – отчет по лабораторной работе.

Тема 8. Нелинейная регрессия (1 ч).

Примерный перечень заданий

Метод Ньютона-Рафсона, метод Ньютона-Гаусса. Логистическая регрессия, метод IRLS.

Обобщенная линейная модель. Экспоненциальное семейство распределений.

Форма контроля – отчет по лабораторной работе.

Тема 11. Методы ансамблирования (1 ч).

Примерный перечень заданий

Бустинг, градиентный бустинг, стохастический градиентный бустинг. Варианты бустинга.

Форма контроля – отчет по лабораторной работе.

Тема 15. Глубокие нейронные сети (1 ч).

Примерный перечень заданий

Полносвязные нейронные сети. Сверточные сети для изображений, ResNet.

Рекуррентные нейронные сети.

Модели внимания и трансформеры.

Форма контроля – отчет по лабораторной работе.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используются:

- *эвристический подход*, который предполагает выбор содержания и способа его организации при подготовке образовательных продуктов (сообщений, докладов, презентаций) по проблемам, решаемым методами машинного обучения и их соотнесения с многообразием решений большинства профессиональных задач и жизненных проблем; творческую самореализацию обучающихся в процессе создания продуктов; индивидуализацию обучения через возможность самостоятельно ставить цели, осуществлять рефлекссию собственной образовательной деятельности;

- *методы и приемы развития критического мышления*, которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения, письма, создания программ; понимании информации как отправного, а не конечного пункта критического мышления.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине «Математические основы машинного обучения» используются современные информационные ресурсы: размещается на образовательном портале комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, материалы текущего контроля и промежуточной аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательного стандарта высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к экзамену, задания, вопросы для самоконтроля и др., список рекомендуемой литературы, информационных ресурсов и др.).

Примерный перечень вопросов к экзамену

1. Постановка основных задач машинного обучения. Объекты и признаки. Типы задач: классификация, регрессия, ранжирование. Методы обучения: с учителем, без учителя, с частичным привлечением учителя, обучение с подкреплением. Основные понятия: модель алгоритмов, функционал качества, минимизация эмпирического риска, обобщающая способность, переобучение.
2. Линейная регрессия. Метод наименьших квадратов. Линейная классификация, аппроксимация пороговой функции потерь.
3. Метод градиентного спуска, стохастического градиентного спуска. Проблема мультиколлинеарности и переобучения.
4. Вероятностная постановка задач классификации. Принцип максимума правдоподобия. Регуляризация, ее вероятностная интерпретация – гауссовский и лапласовский регуляризаторы. Логистическая регрессия.
5. Модель нейрона МакКаллока-Питтса. Функции активации. Нейронные сети как универсальный аппроксиматор. Обучение нейронных сетей – метод обратного распространения ошибки.
6. Обобщенный метрический классификатор. Метод ближайших соседей kNN и его обобщения.
7. Метод окна Парзена. Метод потенциальных функций, связь с линейными моделями.
8. Непараметрическая регрессия. Формула Надарая-Ватсона.
9. Оптимальная разделяющая гиперплоскость, понятие зазора между классами. Случай линейной разделимости. Сведение к задаче квадратичного программирования.
10. Понятие опорных векторов. Функции ядра, спрямляющее пространство, теорема Мерсера. Способы конструктивного построения ядер, примеры.
11. Многомерная линейная регрессия, метод наименьших квадратов, геометрический смысл.
12. Сингулярное разложение, мультиколлинеарность и переобучение. Гребневая регрессия.

13. Метод главных компонент. Понижение размерности, задачи низкоранговых матричных разложений.
14. Метод Ньютона-Рафсона, метод Ньютона-Гаусса. Логистическая регрессия, метод IRLS.
15. Обобщенная линейная модель. Экспоненциальное семейство распределений.
16. Критерии качества классификации: ROC-кривая, точность, полнота, PR-кривая.
17. Внутренние и внешние критерии. Кросс-проверка.
18. Информационный критерий Акаике, байесовский информационный критерий, оценка Вапника-Червоненкиса, VC-размерность.
19. Задача отбора признаков. Полный перебор, жадные алгоритмы, поиск в глубину и ширину.
20. Понятие логической закономерности. Параметрические семейства закономерностей: конъюнкции пороговых правил. Переборные алгоритмы синтеза конъюнкций.
21. Статистический критерий информативности. Тест Фишера.
22. Решающее дерево, жадная стратегия. Алгоритм ID3. Случайный лес. Деревья регрессии.
23. Базовый алгоритм, простое голосование. Бэггинг, случайный лес, AdaBoost.
24. Бустинг, градиентный бустинг, стохастический градиентный бустинг. Варианты бустинга. Стэкинг.
25. Параметрическое оценивание плотности, многомерное нормальное распределение. Выборочные оценки параметров.
26. Непараметрическая оценка плотности, смесь распределений, EM-алгоритм.
27. Оптимальный байесовский классификатор. Генеративные и дискриминативные модели. Наивный байесовский классификатор. Линейный наивный байесовский классификатор в случае экспоненциального семейства распределений.
28. Нормальный дискриминантный анализ. Линейный дискриминант Фишера.
29. Смесь многомерных нормальных распределений. Сеть радиальных базисных функций.
30. Постановка задачи кластеризации. Примеры. Алгоритмы k-means, DBSCAN.
31. Построение дендрограммы. Определение числа кластеров. Другие алгоритмы кластеризации.
32. Полносвязные нейронные сети. Сверточные сети для изображений, ResNet.
33. Рекуррентные нейронные сети.
34. Перенос обучения (transfer learning) и дистилляция нейронных сетей.
35. Модели внимания и трансформеры.
36. Векторные представления текста. Задача машинного перевода.
37. Тематическое моделирование. Латентное размещение Дирихле

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
отсутствует			

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ
ПО ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
теории функций (протокол № ____ от _____ 202_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ

Декан факультета