

Белорусский государственный университет

УТВЕРЖДАЮ
Проректор по учебной работе


А.М. Толстик

12.07.

Регистрационный № УД- 1220 / уч.

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности второй ступени высшего
образования (магистратуры) с углубленной подготовкой специалиста:

1-31 81 09 «Алгоритмы и системы обработки больших объемов
информации».

2017 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 81 09-2014 и учебного плана G31-229/уч. от 26.05.2017.

Составители:

А.В. Колесов – ассистент кафедры дискретной математики и алгоритмики Белорусского государственного университета.

Рекомендована к утверждению:

Кафедрой дискретной математики и алгоритмики Белорусского государственного университета (протокол № 13 от 20 апреля 2017 г.);

Научно-методическим советом Белорусского государственного университета (протокол № 4 от 11 мая 2017 г.).

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная программа по учебной дисциплине «Методы машинного обучения» разработана в соответствии с учебным планом и образовательным стандартом второй ступени высшего образования (магистратуры) с углубленной подготовкой специалиста по специальности 1-31 81 09 «Алгоритмы и системы обработки больших объемов информации».

Учебная дисциплина «Методы машинного обучения» знакомит магистрантов с основными направлениями в теории обучения машин и закладывает необходимую теоретическую базу для понимания и разработки новых алгоритмов машинного обучения.

Основой для изучения учебной дисциплины являются следующие учебные дисциплины первой ступени высшего образования: «Теория вероятностей и математическая статистика», «Вычислительные методы алгебры», «Методы численного анализа» и «Математический анализ».

Цель преподавания учебной дисциплины «Методы машинного обучения»: создание базы для применения современных методов машинного обучения и формирование у магистрантов умения анализировать задачи машинного обучения и осуществлять взвешенный выбор того или иного решения.

При изложении материала учебной дисциплины важно показать многообразие методов машинного обучения и возможность их комбинации для решения поставленной задачи.

Основные задачи, решаемые при изучении учебной дисциплины «Методы машинного обучения»:

- изучение 3 современных языков программирования из различных групп (функциональные статически-типизированные, функциональные динамически-типизированные, объектно-ориентированные динамически-типизированные);

- сравнительный анализ изученных языков программирования;

- использование изученных языков программирования в качестве метаязыка для разработки различных алгоритмов реализации языка программирования.

В результате изучения дисциплины магистрант должен:

знать:

- математические основы теории машинного обучения;
- основные классы алгоритмов машинного обучения и их представителей и их взаимосвязь, достоинства и недостатки;

уметь:

- анализировать, выделять особенности и комбинировать методы машинного обучения;
- применять методы машинного обучения для решения прикладных задач;

владеть:

- программными средствами для разработки алгоритмов машинного обучения, алгоритмами построения искусственных нейронных сетей;
- навыками разработки моделей машинного обучения.

Освоение образовательной программы магистратуры должно обеспечить формирование следующих групп компетенций:

академических компетенций – углубленных научно-теоретических, методологических знаний и исследовательских умений, обеспечивающих разработку научно-исследовательских проектов, инновационной деятельности, непрерывного самообразования (АК-1. Способность к самостоятельной профессиональной деятельности (анализ, сопоставление, систематизация, абстрагирование, моделирование, проверка достоверности данных, принятие решений и др.), готовность генерировать и использовать новые идеи. АК-2. Методологические знания и исследовательские умения, обеспечивающие решение прикладных задач и инновационной деятельности. АК-3. Способность к постоянному самообразованию);

социально-личностных компетенций – личностных качеств и умений следовать социально-культурным и нравственным ценностям; способностей к социальному, межкультурному взаимодействию, критическому мышлению; социальной ответственности, позволяющих решать социально-профессиональные, организационно-управленческие, воспитательные задачи (Магистр должен: СЛК-1. Учитывать социальные и нравственно-этические нормы в социально-профессиональной деятельности. СЛК-2. Быть способным к сотрудничеству и работе в команде. СЛК-3. Владеть коммуникативными способностями для работы в междисциплинарной и международной среде. СЛК-4. Совершенствовать и развивать свой интеллектуальный и общекультурный уровень, добиваться нравственного и физического совершенствования своей личности. СЛК-5. Формировать и аргументировать собственные суждения и профессиональную позицию. СЛК-6. Логично, аргументированно и ясно строить устную и письменную речь, использовать навыки публичной речи, ведения дискуссии и полемики. СЛК-7. Проявлять инициативу и креативность, в том числе в нестандартных ситуациях);

профессиональных компетенций – углубленных знаний по специальным дисциплинам и способностей решать сложные профессиональные задачи, задачи научно-исследовательской и научно-педагогической деятельности, разрабатывать и внедрять инновационные проекты, осуществлять непрерывное профессиональное самосовершенствование (Магистр должен быть способен: ПК-1. Квалифицированно использовать современные достижения по разработке и анализу алгоритмов и современные информационные технологии. ПК-2. Строить математические модели для прикладных задач и проводить теоретическое и экспериментальное их исследование. ПК-3. Разрабатывать эффективные численные алгоритмы и

интегрировать их в компьютерные системы. ПК-4. Обосновывать выбор методов и инструментов для решения прикладных задач. ПК-5. Обосновывать достоверность полученных результатов. ПК-6. Осваивать и реализовывать управленческие инновации в профессиональной деятельности. ПК-7. Формулировать выводы и рекомендации по применению современных достижений науки в инновационной деятельности).

Учебная программа рассчитана на 112 часов, из них 54 аудиторных часа, в том числе 24 лекционных часа и 30 часов лабораторных занятий.

Рекомендуемая форма текущей аттестации – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Введение

История машинного обучения, основные задачи машинного обучения, краткий обзор курса.

Раздел I. Основы машинного обучения

Тема 1.1. Задачи обучения по прецедентам.

Объекты, признаки, ответы, функционал качества. Вероятностная постановка задачи обучения. Переобучение, обобщающая способность. Задачи классификации, восстановления регрессии, ранжирования, кластеризации, поиска ассоциаций.

Тема 1.2. Формальная модель машинного обучения.

Формальная модель обучения: PAC-learnability. Необходимый размер выборки. Agnostic PAC learning. Оптимальный байесовский классификатор. Обучение через равномерную сходимость. Bias-variance tradeoff. VC-размерность. Другие модели обучения.

Раздел II. Основные алгоритмы решения задач классификации и восстановления регрессии

Тема 2.1. Метрические методы машинного обучения.

Обобщённый метрический классификатор. Метод ближайшего соседа. KNN. Метод потенциальных функций. Метод парзеновского окна. Понятие отступа. Понятие эталонного объекта. Проклятие размерности.

Тема 2.2. Байесовские методы машинного обучения.

Оптимальный байесовский классификатор. Задача восстановления плотности распределения. Наивный байесовский классификатор. Непараметрическая оценка плотности. Нормальный дискриминантный анализ. Разделение смеси распределений. EM-алгоритм.

Тема 2.3. Линейные методы машинного обучения и их обобщения.

Аппроксимация эмпирического риска. Линейный классификатор и линейная регрессия. Метод стохастического градиентного спуска. Логистическая регрессия. Метод опорных векторов. Kernel trick.

Тема 2.4. Решающие деревья.

Решающие деревья. Случайный лес. Бустинг и бэггинг. Примеры решающих деревьев. Информативность. Алгоритмы построения решающих деревьев.

Раздел III. Избранные главы машинного обучения

Тема 3.1. Визуализация и кластеризация.

Качество кластеризации. Статистические, эвристические методы кластеризации. Иерархическая кластеризация. Сети Кохонена. t-SNE. PCA.

Тема 3.2. Искусственные нейронные сети.

Модель нейрона. Многослойный перцептрон. Метод обратного распространения ошибки. Свёрточные нейронные сети. Рекуррентные нейронные сети.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

№ п/п	Название раздела, темы	Количество часов				Количество часов УСР	Форма контроля знаний
		Аудиторные					
		Лекции	Пр. и сем. зан.	Лаб. зан.	Иное		
1	Основы машинного обучения	4		4			
1.1	Задачи обучения по прецедентам	2					Устный опрос
	<i>Лабораторная работа 1. Основы анализа данных на языке Python</i>			2			Защита лабораторной работы 1
1.2	Формальная модель машинного обучения	2					Устный опрос
	<i>Лабораторная работа 2. Модели машинного обучения</i>			2			Защита лабораторной работы 2
2	Основные алгоритмы решения задач классификации и восстановления регрессии	16		20			
2.1	Метрические методы машинного обучения	4					Устный опрос
	<i>Лабораторная работа 3. Метод ближайших соседей</i>			4			Защита лабораторной работы 3
2.2	Байесовские методы машинного обучения	4					Устный опрос
	<i>Лабораторная работа 4. Методы восстановления плотности и байесовские классификаторы</i>			6			Защита лабораторной работы 4
2.3	Линейные методы машинного обучения и их обобщения	4					Устный опрос
	<i>Лабораторная работа 5. Логистическая регрессия и метод опорных векторов</i>			4			Защита лабораторной работы 5
2.4	Решающие деревья	4					Устный опрос

	<i>Лабораторная работа 6. Бустинг над решающими деревьями</i>			6			Защита лабораторной работы 6
3.	Избранные главы машинного обучения	4		6			
3.1	Визуализация и кластеризация	2					Устный опрос
	<i>Лабораторная работа 7. t-SNE и PCA</i>			2			Защита лабораторной работы 7
3.2	Искусственные нейронные сети	2					Устный опрос
	<i>Лабораторная работа 8. Нейронные сети</i>			4			Защита лабораторной работы 8
ИТОГО		24		30			

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Рекомендуемая литература

Основная

1. Shalev-Shwartz Shai, Ben-David Shai. Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014. — 409 p. — ISBN-13: 978-1107057135.
2. Bishop, C.M. Pattern Recognition and Machine Learning. — Springer, 2006. — 738 p.
3. Вапник, В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979. — 448 с.
4. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd edition. — Springer, 2009. — 533 p.

Дополнительная

1. Cristianini, N. & Shawe-Taylor, J. (2000), An Introduction to Support Vector Machines, Cambridge University Press.
2. Kearns, M. & Vazirani, U. (1994), An Introduction to Computational Learning Theory, MIT Press.
3. Nemirovski, A. & Yudin, D. (1978), Problem complexity and method efficiency in optimization, Nauka Publishers, Moscow.

Рекомендации по контролю качества усвоения знаний и проведению аттестации

На лекционных занятиях по учебной дисциплине «Методы машинного обучения» рекомендуется использование элементов проблемного обучения: проблемное изложение некоторых аспектов, использование частично-поискового метода.

Перечни используемых средств диагностики результатов учебной деятельности

Для аттестации обучающихся на соответствие их персональных достижений поэтапным и конечным требованиям образовательной программы создаются фонды оценочных средств, включающие типовые задания, контрольные работы и тесты. Оценочными средствами предусматривается оценка способности обучающихся к творческой деятельности, их готовность вести поиск решения новых задач, связанных с недостаточностью конкретных специальных знаний и отсутствием общепринятых алгоритмов.

Для диагностики компетенций в рамках учебной дисциплины рекомендуется использовать следующие формы:

1. Устная форма: опросы, устная защита лабораторных работ.
2. Письменная форма: отчеты по лабораторным работам, оценивание на основе модульно-рейтинговой системы.

Контрольные мероприятия проводятся в соответствии с учебно-методической картой дисциплины. В случае неявки на контрольное мероприятие по уважительной причине студент вправе по согласованию с преподавателем выполнить его в дополнительное время. Для студентов, получивших неудовлетворительные оценки за контрольные мероприятия, либо не явившихся по неуважительной причине, по согласованию с преподавателем и с разрешения заведующего кафедрой мероприятие может быть проведено повторно.

Оценка текущей успеваемости рассчитывается как среднее за отчеты по домашним практическим упражнениям и лабораторные работы.

Итоговая аттестация предусматривает проведение зачета. При этом рекомендуется использовать оценивание успеваемости на основе модульно-рейтинговой системы.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ

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)
Математический анализ	Высшей математики	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 20.04.2017 г
Теория вероятностей и математическая статистика	Теории вероятностей и математической статистики	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 20.04.2017 г
Вычислительные методы алгебры	Вычислительной математики	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 20.04.2017 г
Методы численного анализа	Вычислительной математики	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 20.04.2017 г

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

№№ Пп	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры дискретной математики и алгоритмики (протокол № ____ от _____ 201_ г.)
Заведующий кафедрой

(ученая степень, звание)

(подпись)

(И.О. Фамилия)

УТВЕРЖДАЮ
Декан факультета

(ученая степень, звание)

(подпись)

(И.О.Фамилия)