

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.ЛОМОНОСОВА»
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ



УТВЕРЖДАЮ

Декан факультета ВМК МГУ,

Академик

«14» сентября 2022 г.

/И.А. Соколов/

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

**Методы машинного обучения и поиск достоверных закономерностей в
данных**

Machine learning methods and search for valid regularities in data

Программа (программы) подготовки научных и научно-педагогических кадров в аспирантуре

Рабочая программа дисциплины разработана в соответствии с Приказом Ректора МГУ №1216 от 24 ноября 2021 года «Об утверждении Требований к основным программам подготовки научных и научно-педагогических кадров в аспирантуре, самостоятельно устанавливаемых Московским государственным университетом имени М.В. Ломоносова»

1. Краткая аннотация:

Название дисциплины Методы машинного обучения и поиск достоверных закономерностей в данных

Цель изучения дисциплины – Данный курс, дополняет основной университетский поточный курс, что позволяет существенно расширить знания аспиранта. Курс посвящен изложению теоретических основ известных методов машинного обучения и интеллектуального анализа данных. В нем рассматриваются статистические, комбинаторно-логические, нейросетевые, ансамблевые методы машинного обучения, метод опорных векторов. Обсуждаются преимущества и недостатки различных моделей, а также теоретически обоснованные способы повышения эффективности обучения. Большое внимание уделяется методам оценки точности и эффективности полученных решений. Излагаются известные методы кластерного анализа и поиска закономерностей в данных. Рассматриваются современные методы оценки статистической значимости выявленных закономерностей. Обсуждаются также распространённые ошибки, возникающие при использовании методов машинного обучения и интеллектуального анализа данных.

2. Уровень высшего образования – аспирантура

3. Научная специальность 1.2.1., 1.2.2., 1.2.3., 1.1.2., 1.1.4., 1.1.5., 1.1.6., 2.3.5., 2.3.6., отрасль науки: Физико-математические науки,

Научная специальность 1.2.2., отрасль науки: Технические науки

4. Место дисциплины (модуля) в структуре Программы аспирантуры – элективный курс.

5. Объем дисциплины (модуля) составляет 2 зачетные единицы, всего 72 часа, из которых 28 часа составляет контактная работа аспиранта с преподавателем, 44 часа составляет самостоятельная работа учащегося.

6. Входные требования для освоения дисциплины (модуля), предварительные условия.

На предыдущих уровнях высшего образования должны быть освоены общие курсы:

1. Математический анализ
2. Линейная алгебра
3. Дискретная математика
4. Теория вероятностей и математическая статистика
5. Оптимальное управление.

7. Содержание дисциплины (модуля), структурированное по темам

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы						Самостоятельная работа обучающегося, часы		
		из них						из них		
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости, промежуточной аттестации	Всего	Выполнение домашних заданий	Подготовка к коллоквиумам	Всего
Тема 1.Цели и задачи машинного обучения (МО). Способы оценки эффективности использования МО. Обсуждается основная цель машинного обучения – достижение максимальной обобщающей способности. Рассматривают основные способы оценки	6	2		-	-		2	4	-	4

эффективности в смысле указанной цели.										
Тема 2. Современные методы регрессионного анализа, включая подходы, основанные на регуляризации по Тихонову Рассказывается о методах линейной регрессии, использующих регуляризаторы различного типа, включая гребневую регрессию, Лассо, эластичную сеть.	10	2	-	-	-	-	2	8	-	8
Тема 3. Статистические методы решения задач классификации Рассказывается об основных статистических методах решения задач классификации	10	2	-	-	-	-	2	8	-	8

Тема 4. Метод опорных векторов в задачах классификации с учителем Обсуждаются теоретические основы и особенности применения метода опорных векторов.	10	2		-	-		2	8	-	8
Тема 5. Метод решающих деревьев в задачах классификации с учителем Обсуждаются теоретические основы метода решающих деревьев	6	2					2	4		4
Тема 6. Ансамблевые методы решения задач классификации с учителем. Теоретически обосновывается использование ансамблей классификаторов. Рассказывается о	14	4					4	12		12

способах генерации ансамблей и вычисления коллективных решений.										
Тема 7 Нейросетевые методы решения задач классификации с учителем. Рассказывается об архитектуре и способах обучения нейросетевых моделей.	16	6					6	10		10
Тема 8 Байесовские сети Даётся краткое введение в байесовские сети	6	2					2	4		4
Тема 9 Методы интеллектуального анализа данных Рассказывается о методах кластерного анализа. Рассказывается также о способах поиска статистически значимых закономерностей.	8	4				4	8			

Аттестация (экзамен)	2									
Итого	72						28			44

8. Образовательные технологии.

При чтении лекций в качестве материала, иллюстрирующего возможности математического моделирования в различных ситуациях, активно используются примеры из практики обработки данных в процессе исследований в предметной области. Информационные и интерактивные технологии используются при обсуждении проблемных и неоднозначных вопросов, требующих выработки решения в ситуации неопределенности.

9. Учебно-методические материалы для самостоятельной работы по дисциплине (модулю):

Самостоятельная работа учащихся состоит в изучении лекционного материала, учебно-методической литературы, подготовки к текущему контролю и промежуточной аттестации.

Литература для самостоятельной работы студентов в соответствии с тематическим планом .

Тема 1 «Цели и задачи машинного обучения (МО). Способы оценки эффективности использования МО»

1. Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,» Москва, Фазис, (2006).

Тема 2 «Современные методы регрессионного анализа, включая подходы, основанные на регуляризации по Тихонову»

1. T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008

Тема 3 «Статистические методы решения задач классификации с учителем»

- 1 Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,» Москва, Фазис, (2006).
- 2 T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008

Тема 4 «Метод опорных векторов в задачах классификации с учителем»

- 1 Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,» Москва, Фазис, (2006).
- 2 T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008

- 3 C.J.C. Burges. “A Tutorial on Support Vector Machines for Pattern Recognition,” Data Mining and Knowledge Discovery 2, 121-167(1998)

Тема 5 «. Метод решающих деревьев в задачах классификации с учителем»

1. *Quinlan, J. R. [Induction of Decision Trees](#) (англ.) // Machine Learning. — Kluwer Academic Publishers, 1986. — No. 1. — P. 81—106*
2. Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,” Москва, Фазис, (2006).
3. T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008

Тема 6 «Ансамблевые методы решения задач классификации с учителем»

1. Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,” Москва, Фазис, (2006).
2. T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008

Тема 7 «Нейросетевые методы решения задач классификации с учителем.

»

1. Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,” Москва, Фазис, (2006).
2. T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008
3. *Schmidhuber, J. Deep Learning in Neural Networks: An Overview (неопр.) // Neural Networks. — 2015. — Т. 61. — С. 85—117.*

Тема 8 «Байесовские сети».

1. *Judea Pearl. Causality: Models, Reasoning, and Inference. — 2-nd Edition. — Cambridge University Press, 2009. — 464 p. — [ISBN 9780521895606](#).*

Тема 9 «Методы интеллектуального анализа данных».

1. Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько, «Распознавание». Математические методы. Программная система. Практические применения,” Москва, Фазис, (2006).
 2. T.Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Second Edition Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, 2008
 3. Миркин, Б. Г. Методы кластер-анализа для поддержки принятия решений: обзор : препринт WP7/2011/03 [Текст] / Б. Г. Миркин ; Национальный исследовательский университет «Высшая школа экономики». — М. : Изд. дом Национального исследовательского университета «Высшая школа экономики», 2011. — 88 с. — 150 экз.
- Описание материально-технической базы.
Занятия проводятся в аудитории, оснащенной мультимедийным экраном

11. Язык преподавания – русский

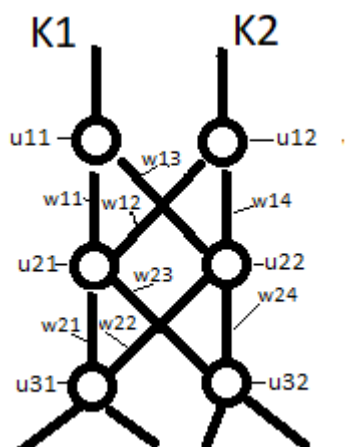
12. Преподаватели:

Степень, должность ФИО., e-mail, тел.: Сенько Олег Валентинович senkoov@mail.ru, 4959395135

Фонды оценочных средств, необходимые для оценки результатов обучения/

Примеры домашних заданий

1)



На рисунок

На рисунке изображена верхняя часть многослойной нейронной сети.

Заданы значения весов связей между нейронами

$w_{11}=0.73$ $w_{21}=1.18$

$w_{12}=-1.23$ $w_{22}=0.44$

$w_{13}=1.13$ $w_{23}=0.96$

$w_{14}=0.46$ $w_{24}=-0.76$

Для обучения нейронной сети используется метод обратного распространения ошибки BP с константой скорости обучения $d=0.05$

Известно, при для некоторого поданного на вход сети объекта s из класса K1 были вычислены следующие значения для сигналов предпоследнего внутреннего слоя $u_{31}=0.43$ $u_{32}=0.56$

Задание

- 1) Вычислить u_{21} , u_{22} , u_{11} , u_{12}
- 2) Как изменится вес w_{11} после коррекции весов с помощью BP,
- 3) Как изменится вес w_{22} после коррекции весов с помощью BP,

2) Имеется 2 класса объектов K_1 и K_2 , описываемых непрерывным признаком X_1 и категориальным признаком $X_2 \in \{a, b\}$

В классе K_1 признак X_1 подчиняется нормальному распределению с математическим ожиданием $\mu = -1$ и стандартным отклонением $\sigma = 2$, признак X_2 подчиняется распределению: $P(a|K_1) = 0.2$, $P(b|K_1) = 0.8$.

В классе K_2 признак X_1 подчиняется нормальному распределению с математическим ожиданием $\mu = 1$ и стандартным отклонением $\sigma = 1$, признак X_2 подчиняется распределению: $P(a|K_2) = 0.6$, $P(b|K_2) = 0.4$.

Априорные вероятности $P(K_1) = 0.4$, $P(K_2) = 0.6$.

- 1) Определить множество значений признака X_1 , при которых наивный байесовский классификатор (НБК) будет относить объект в класс K_1 при условии,

а) $X_2 = a$

б) $X_2 = b$

- 3 Определить, при каком условии на $P(a|K_1)$ и $P(a|K_2)$ множество значений признака X_1 , при которых НБК относит объект в класс K_2 , является пустым

2)

Банк испытал технологию машинного обучения для решения задачи кредитного скоринга

Известно, что доход банка от одного добросовестного заёмщика 43000 ус.ед.

Потери от недобросовестного заёмщика 122000 ус.ед. . Доля недобросовестных заёмщиков -11%.

Связь Чувствительности (TPR) и Ложной тревоги (FPR) представлена в таблице

TPR	FPR
0.05	0.0001
0.08	0.001
0.16	0.04
0.22	0.08
0.34	0.11
0.47	0.17
0.56	0.23
0.67	0.33
0.78	0.45
0.87	0.56
0.95	0.65
0.99	0.77

Определить, какой дополнительный доход в пересчёте на одну поданную заявку даст разработанная технология.