

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»

ФИЗИКО-ТЕХНИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра компьютерных технологий

УТВЕРЖДАЮ:

Проректор по научно-методической
и учебной работе

Е.И. Скафа

«28» июня 2017 г.



Рабочая программа учебной дисциплины

«МАШИННОЕ ОБУЧЕНИЕ»

Направления подготовки:	09.04.01 Информатика и вычислительная техника
Магистерская программа:	Информатика и вычислительная техника
Программа подготовки:	академическая магистратура
Квалификация:	магистр
Форма обучения:	очная

Донецк 2017



УТВЕРЖДАЮ:

Декан физико-технического факультета

Н.Г. Малюк

«23» июня 2017 г.

МП

Программа составлена с учетом Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.04.01 Информатика и вычислительная техника, утвержденного приказом Министерства образования и науки Российской Федерации от 30 октября 2014 г. № 1420.

Программа учебной дисциплины «Машинное обучение» составлена на основе ГОС ВПО по направлению подготовки 09.04.01 Информатика и вычислительная техника, утвержденного приказом Министерства образования и науки ДНР от «25» декабря 2015 г. № 946, зарегистрированного в Министерстве юстиции ДНР от 01 февраля 2016 г. № 948, «Положения об организации учебного процесса в образовательных организациях высшего профессионального образования Донецкой Народной Республики», утвержденного приказом Министерства образования и науки ДНР от 07 августа 2015 г. № 380 (с изменениями и дополнениями от 30 октября 2015 г. № 750), учебного плана по направлению подготовки 09.04.01 Информатика и вычислительная техника (Профиль: Информатика и вычислительная техника), утвержденного Ученым Советом Университета от 31.03.2017 г., протокол № 3 и основной образовательной программы, утвержденной приказом ректора (№ 77/05 от 06.05 2017 г.).

Разработчик:

канд. техн. наук,

доцент кафедры компьютерных технологий

Т.В. Ермоленко

Программа учебной дисциплины утверждена на заседании кафедры компьютерных технологий

Протокол № 17 от «04» мая 2017 г.

Заведующий кафедрой

Т.В. Ермоленко

Программа учебной дисциплины одобрена учебно-методической комиссией физико-технического факультета

Протокол № 5 от «24» мая 2017 г.

Председатель учебно-методической комиссии факультета

В.Н. Котенко

1. Область применения и место дисциплины в учебном процессе

Учебная дисциплина «Машинное обучение» относится к вариативной части блока «Дисциплины (модули)» и состоит из двух содержательных модулей: модуль 1 – «Методы классификации и регрессии», модуль 2 – «Ансамбли решающих правил. Критерии выбора моделей и методы отбора признаков».

Дисциплина реализуется на физико-техническом факультете кафедрой компьютерных технологий.

Этот курс опирается на математическую подготовку студентов, полученную при изучении дисциплин бакалавриата: «Математика», «Теория вероятности, математическая статистика», «Математическая логика», «Современные информационные системы и технологии», на знания технологий разработки современного программного обеспечения, полученные при изучении дисциплин бакалавриата: «Технологии разработки программного обеспечения» «Объектно-ориентированное программирование», а также на знания технологий искусственного интеллекта, полученные в результате освоения дисциплин магистратуры «Технологии извлечения знаний» и «Интеллектуальные системы», закладывает фундамент научно-прикладной подготовки будущих исследователей в области инженерии знаний.

Полученные знания используются студентами во время выполнения научно-исследовательской работы и при написании магистерской диссертации.

2. Структура дисциплины

<i>Характеристика учебной дисциплины</i>		
Направление подготовки	09.04.01 Информатика и вычислительная техника	
Магистерская программа	Информатика и вычислительная техника	
Программа подготовки	академическая магистратура	
Квалификация	магистр	
Количество содержательных модулей	2	
Дисциплина базовой / вариативной части образовательной программы	дисциплина вариативной части	
Формы контроля	1 модульный контроль, 1 экзамен	
Показатели	очная форма обучения	заочная форма обучения
Количество зачетных единиц (кредитов)	4,5	
Год подготовки	2	
Семестр	3	
Количество часов	162	
- лекционных	18	
- практических, семинарских	36	
- лабораторных	36	
- самостоятельной работы	72	
в т.ч. индивидуальное задание		
Недельное количество часов,	9	
в т.ч. аудиторных	5	

3. Описание дисциплины

Цели и задачи.

Цель – формирование у магистрантов знаний основных математических методов и алгоритмов машинного обучения, развитие интуицию для лучшего понимания основных идей, лежащих за этими методами, привить навыки работы с программным обеспечением, реализующим алгоритмы машинного обучения, и умение применять полученные знания при проектировании и реализации интеллектуальных информационных систем.

Задачи – освоение логических, метрических и вероятностных методов классификации и регрессии, алгоритмов построения разделяющих гиперплоскостей и нейросетей, критериев выбора моделей и методов отбора признаков, методов построения композиций классификаторов.

Требования к результатам освоения дисциплины:

Процесс изучения дисциплины «Машинное обучение» направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО РФ по направлению подготовки 09.04.01 Информатика и вычислительная техника и основной образовательной программы высшего профессионального образования направления подготовки 09.04.01 Информатика и вычислительная техника (магистерская программа: Информатика и вычислительная техника)

а) общекультурных (ОК):

- способностью к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности (ОК-3);
- способностью заниматься научными исследованиями (ОК-4);
- способностью самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности (ОК-7);
- умение оформлять отчеты о проведенной научно-исследовательской работе и подготавливать публикации по результатам исследования (ОК-9).

б) общепрофессиональных (ОПК):

- способностью воспринимать математические, естественнонаучные, социально-экономические и профессиональные знания, умением самостоятельно приобретать, развивать и применять их для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте (ОПК-1);
- культурой мышления, способностью выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных (ОПК-2);
- владением методами и средствами получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях (ОПК-5);
- способностью анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями (ОПК-6).

в) профессиональных (ПК):

научно-исследовательская деятельность:

- владением существующими методами и алгоритмами решения задач распознавания и обработки данных (ПК-4);
- применением перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий (ПК-7);

производственно-технологическая деятельность:

— способностью к созданию программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов (ПК-15).

В результате изучения учебной дисциплины студент должен**Знать:**

- постановку основных задач машинного обучения;
- возможности алгоритмов машинного обучения и способы их применения для решения практических задач управления и распознавания образов;
- логические, метрические и вероятностные методы классификации и регрессии;
- основные понятия и принципы работы искусственных нейронных сетей;
- композиции классификаторов, методы бустинга;
- критерии выбора моделей и методы отбора признаков;

Уметь:

- выполнять грамотную постановку задач, возникающих в практической деятельности, для их решения с помощью методов машинного обучения;
- обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи;
- проводить анализ работы методов машинного обучения с выявлением их сильных и слабых сторон;
- анализировать результаты обучения алгоритма машинного обучения, предлагать пути повышения его точности;
- программно реализовывать алгоритмы машинного обучения;

Владеть:

- базовым инструментарием машинного обучения;
- навыками построения различного рода классификаторов.

4. Содержание дисциплины и формы организации учебного процесса

Порядковый номер и тема	Краткое содержание темы
1	2
Содержательный модуль 1. Инженерия знаний. Системы, основанные на знаниях	
Тема 1. Основные понятия и примеры прикладных задач.	Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал. Типы задач. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.

1	2
<p>Тема 2. Методы классификации и регрессии.</p>	<p>Метрические и логические методы классификации и регрессии: метод эталонных образов, метод ближайших соседей kNN и его обобщения; подбор числа k по критерию скользящего контроля; логическая закономерность; параметрические семейства закономерностей; переборные алгоритмы синтеза конъюнкций; решающее дерево; вывод критериев ветвления; энтропийный критерий, критерий Джини. Линейный классификатор: разделяющая поверхность; метод решающих функций и опорных векторов; линейные модели регрессии и классификации; метод наименьших квадратов; полиномиальная регрессия. Байесовская теория классификации: наивный байесовский классификатор; непараметрические и параметрические методы оценивания плотности; смеси. Нейронные сети: модель биологического нейрона; функции активации; алгоритм обратного распространения ошибок; свёрточные сети, рекуррентные сети.</p>
<p>Тема 3. Прогнозирование временных рядов.</p>	<p>Задача прогнозирования временных рядов. Примеры приложений. Экспоненциальное скользящее среднее. Адаптивная авторегрессионная модель. Адаптивная селективная модель. Адаптивная композиция моделей.</p>
<p>Тема 4. Нечеткие знания.</p>	<p>Нечеткие множества и отношения. Теория приближенных рассуждений. Мягкая экспертная система.</p>
<p align="center">Содержательный модуль 2. Ансамбли решающих правил. Критерии выбора моделей и методы отбора признаков</p>	
<p>Тема 4. Линейные композиции, бустинг.</p>	<p>Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция. Взвешенное голосование. Алгоритм AdaBoost. Экспоненциальная аппроксимация пороговой функции потерь. Процесс последовательного обучения базовых алгоритмов. Теорема о сходимости бустинга. Базовые алгоритмы в бустинге. Алгоритм AnyBoost. Алгоритм ComBoost.</p>
<p>Тема 5. Критерии выбора моделей и методы отбора признаков.</p>	<p>Внутренние и внешние критерии. Эмпирические, статистические и аналитические критерии. Сложность задачи отбора признаков. Полный перебор. Метод добавления и удаления, шаговая регрессия. Поиск в глубину, метод ветвей и границ. Анализ главных компонент и факторный анализ.</p>

Тематический план

Содержательный модуль 1												
Названия содержательных модулей и тем	Количество часов											
	Очная форма обучения						Заочная форма обучения					
	всего	в т.ч.					всего	в т.ч.				
		лекции	практические	лабораторные	самостоятельная работа	индивидуальная работа		лекции	практические	лабораторные	самостоятельная работа	индивидуальная работа
Тема 1. Основные понятия и примеры прикладных задач.	4	2			2							
Тема 2. Методы классификации и регрессии.	66	6	16	16	28							
Тема 3. Прогнозирование временных рядов.	20	2	4	4	10							
Итого по содержательному модулю 1	90	10	20	20	40							
Содержательный модуль 2												
Тема 4. Линейные композиции, бустинг.	36	4	8	8	16							
Тема 5. Критерии выбора моделей и методы отбора признаков.	36	4	8	8	16							
Итого по содержательному модулю 2	72	8	16	16	32							
Всего часов	162	18	36	36	72							

Курс дисциплины «Машинное обучение» предусматривает следующие **формы организации учебного процесса**:

- лекции;
- практические занятия;
- лабораторные занятия;
- самостоятельная работа студента.

По источнику передачи и восприятия учебной информации используются словесные (лекция, беседа), наглядные (слайды, иллюстрации, коды программ), практические (исследования, упражнения, лабораторные работы) методы.

По характеру познавательной деятельности студентов используются объяснительно-иллюстративные и репродуктивные методы, проблемное преподавание, частично-поисковый и исследовательский методы.

В зависимости от основной дидактической цели и задач используются методы устного изложения знаний, закрепление учебного материала, самостоятельной работы студентов по осмыслению и усвоению нового материала, работы по применению знаний на практике и выработке умений и навыков, проверки и оценки знаний, умений и навыков.

Используются следующие методы контроля:

- устный контроль (экспресс-опрос на лекциях);
- защита лабораторных работ;
- модульная контрольная работа.
- итоговый контроль (экзаменационные билеты).

5. Методические рекомендации для проведения лекционных, практических и лабораторных занятий

ТЕМЫ ЛЕКЦИОННЫХ ЗАНЯТИЙ

<i>№ п/п</i>	<i>Название темы</i>	<i>Количество часов</i>
1	Основные понятия и примеры прикладных задач.	2
2	Логические методы классификации и регрессии.	2
3	Вероятностные методы классификации и регрессии.	2
4	Искусственные нейронные сети.	2
5	Прогнозирование временных рядов.	2
6	Алгоритмы Adaboost, Anyboost.	2
7	Голосование по большинству и старшинству.	2
8	Отбор информативных признаков.	2
9	Синтез информативных признаков.	2
	ВСЕГО	18

ТЕМЫ ЛАБОРАТОРНЫХ ЗАНЯТИЙ

<i>№ п/п</i>	<i>Название темы</i>	<i>Количество часов</i>
1	Построение модели парной и множественной линейной регрессии средствами языка R.	8
2	Построение наивного байесовского классификатора.	8
3	Анализ временных рядов средствами языка R.	4
4	Анализ данных с помощью метода главных компонент средствами языка R.	16
	ВСЕГО	36

ТЕМЫ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

<i>№ п/п</i>	<i>Название темы</i>	<i>Количество Часов</i>
1	2	3
1	Метод ближайшего соседа и его модификации.	2
2	Отбор эталонных объектов.	2
3	Критерии информативности. Методы поиска информативных закономерностей.	2
4	Решающие списки. Решающие деревья.	2
5	Разделяющая поверхность. Метод решающих функций и опорных векторов	2

1	2	3
6	Модели регрессии. Метод наименьших квадратов.	2
7	Метод потенциальных функций и его связь с линейной моделью классификации.	2
8	Параметрические и непараметрические методы оценивания плотности вероятностей.	2
9	Анализ тренда и сезонности временного ряда.	2
10	Прогнозирование временных рядов.	2
11	Алгоритм Adaboost.	2
12	Алгоритм Anybost.	2
13	Голосование по большинству.	2
14	Голосование по старшинству.	2
15	Критерий средней ошибки и скользящего контроля для выбора модели классификации и регрессии.	2
16	Оценки обобщающей способности классификаторов.	2
17	Алгоритм полного перебора для отбора признаков.	2
18	Последовательное добавление и удаление признаков.	2
	ВСЕГО	36

6. Методические рекомендации по организации самостоятельной работы студентов

- Самостоятельная работа студентов по курсу «Машинное обучение» предусматривает:
- систематическое посещение лекционных занятий, ведение конспекта лекций;
 - повседневное изучение лекционного материала и содержания технической литературы, рекомендуемые этой программой и рабочим учебным планом;
 - добросовестную подготовку к лабораторным и практическим занятиям;
 - своевременное и качественное оформление отчётов по лабораторным работам.
 - самостоятельную разработку алгоритмов и текстов программ лабораторных работ.

ОРГАНИЗАЦИЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

№ п/п	Название темы	Количество часов
1	2	3
1	Изучение темы: Обучение по прецедентам. Обучение без учителя. Примеры практических задач	2
2	Изучение темы: Метрические методы классификации и регрессии. В том числе вопросов, не освещаемых на лекциях: – метод окна Парзена с постоянной и переменной шириной окна; – непараметрическая регрессия. Локально взвешенный метод наименьших квадратов. Ядерное сглаживание; – робастная непараметрическая регрессия. Алгоритм LOWESS.	4
3	Изучение темы: Логические методы классификации и регрессии.	4
	Подготовка к лабораторной работе №1: Изучение возможностей языка R для построения моделей парной и множественной линейной регрессии и проведение их анализа.	4

1	2	3
4	Изучение темы: Вероятностные методы классификации и регрессии. В том числе вопросов, не освещаемых на лекциях: – непараметрическое оценивание плотности: ядерная оценка плотности Парзена-Розенблатта. – параметрическое оценивание плотности: нормальный дискриминантный анализ; – выбор числа компонентов смеси. Пошаговая стратегия. Априорное распределение Дирихле; – смесь многомерных нормальных распределений. Сеть радиальных базисных функций (RBF) и применение ЕМ-алгоритма для её настройки; – сравнение RBF-сети и SVM с гауссовским ядром.	6
	Подготовка к лабораторной работе №2: Изучение байесовского подхода к классификации.	4
5	Изучение темы: Нейронные сети. В том числе вопросов, не освещаемых на лекциях: – быстрые методы стохастического градиента; – методы постепенного усложнения сети, оптимальное прореживание нейронных сетей; – функции активации ReLU и PReLU. Проблема «паралича» сети. – обучение рекуррентных сетей: Backpropagation Through Time (BPTT).	6
6	Изучение темы: Прогнозирование временных рядов. В том числе вопросов, не освещаемых на лекциях: – прогнозирование временных рядов с помощью рекуррентной сети LSTM	4
7	Подготовка к лабораторной работе №3: Изучение возможностей языка R для прогнозирования временных рядов.	6
8	Изучение темы: Ансамбли моделей классификаторов. В том числе вопросов, не освещаемых на лекциях: – бэггинг и метод случайных подпространств; – случайный лес; – градиентный бустинг; – алгоритм LogitBoost; – смесь алгоритмов.	8
9	Изучение темы: Методы голосования. В том числе вопросов, не освещаемых на лекциях: – простое голосование: алгоритм ComBoost.	8
	Изучение темы: Критерии выбора модели. В том числе вопросов, не освещаемых на лекциях: – критерий непротиворечивости моделей; – критерий регуляризации.	4
	Изучение темы: Отбор информативных признаков. В том числе вопросов, не освещаемых на лекциях: – отбор признаков методом стохастического поиска.	4
10	Подготовка к лабораторной работе №4: Изучение метода анализа главных компонент. Изучение возможностей языка R для проведения анализа данных с помощью анализа главных компонент.	8
	ВСЕГО	72

7. Индивидуальные задания.

Индивидуальные задания не предусмотрены.

8. Контрольные вопросы к промежуточной аттестации

1. Функциональная и вероятностная постановка задачи обучения по прецедентам.
2. Типы признаков. Понятие модели алгоритмов и метода обучения. Обобщающая способность метода. Эмпирические оценки обобщающей способности.
3. Функция потерь и функционал качества. Наиболее употребительные функции потерь.
4. Суть метода ближайшего соседа. Какими недостатками он обладает?
5. Алгоритм k ближайших соседей и его модификация – k взвешенных ближайших соседей. Какие недостатки простейших метрических алгоритмов типа kNN?
6. Отступ объекта относительно алгоритма. Типы объектов в порядке убывания отступа. Алгоритм STOLP для отбора эталонных объектов.
7. Критерии информативности, позволяющие называть предикаты закономерностями. Приведите эвристическое, статистическое, энтропийное определения информативности.
8. Бинаризация количественных признаков. Разбиение диапазона значений признака на информативные зоны.
9. Алгоритмы синтеза конъюнкций.
10. Решающие списки, разновидности решающих списков. Жадный алгоритм их построения.
11. Решающие деревья, синтез решающих деревьев, критерии расщепления. Преимущества метода деревьев решений.
12. Основные этапы алгоритма построения решающего дерева ID3, его преимущества и недостатки. Алгоритм CART. Редукция решающих деревьев.
13. Решающая функция и дискриминантная поверхность. Каким свойством обладает гиперплоскость, построенная методом SVM? Зачем в SVM используются ядра?
14. Общий вид регрессионной модели. Модель множественной линейной регрессии. Могут ли зависимости в регрессионной модели быть нелинейными?
15. Подход для нахождения оптимальных оценок параметров регрессионной модели. Оценки качества регрессионной модели.
16. Общая идея метода потенциальных функций. Каковы основные этапы алгоритма метода потенциальных функций?
17. Байесовский подход к распознаванию образов.
18. Параметрические и непараметрические методы оценивания плотности вероятностей. Смеси.
19. Модель нейрона, ее ограничения. Модель многослойного персептрона.
20. Нейронная сеть Хопфилда, ее ограничения.
21. Основные правила обучения нейросетей. Алгоритм обратного распространения ошибки.
22. Общая архитектура сверточных нейросетей. Двумерная свёртка. Техники свёртки, «фильтр» и «ядро» свёртки. Слой пулинга.
23. Глубокие нейросети. Подход к обучению глубоких нейросетей.
24. Схема однослойной рекуррентной сети. Отличие рекуррентной сети Элмана от предыдущих концепций рекуррентных сетей.
25. Общая архитектура рекуррентной сети. Особенности архитектуры LSTM-сети, преимущества LSTM-сети.
26. Алгоритм обучения нейросети методами Хебба.
27. Принцип работы сети Кохонена.

28. Составляющие временного ряда. Модель временного ряда. Типы моделей временного ряда. Трансформация данных ряда, виды трансформации данных.

29. Стационарный временной ряд, тесты для определения стационарности временного ряда. Единичный корень.

30. Коэффициент автокорреляции. Его статистическая оценка. Коррелограмма. Частичная автокорреляция.

31. Наиболее часто используемые на практике критерии проверки «наличия-отсутствия» тренда?

32. Критерии оценки информативности модели ряда. Этапы алгоритма анализа временного ряда.

33. Модель AR(p). Модель MA(p). Модель ARIMA. Определение порядков моделей.

9. Образец модульного контроля

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»

Физико-технический факультет

Направление подготовки: **09.04.01 Информатика и вычислительная техника**
 Магистерская программа: **Информатика и вычислительная техника**
 Программа подготовки: **академическая магистратура**
 Семестр: **3**
 Учебная дисциплина: **Машинное обучение**

МОДУЛЬНАЯ КОНТРОЛЬНАЯ РАБОТА

ВАРИАНТ №1

1. Продукционные системы. Структура продукционной системы и стратегии вывода.
2. Композиционное правило вывода в нечеткой логике.

Утверждено на заседании кафедрой компьютерных технологий,
 протокол № ____ от “__” _____ 20__ г.

Зав. кафедрой _____

Преподаватель _____

Критерии оценивания модульного контроля

<i>Номер задания</i>	<i>Количество баллов</i>
Задание 1	5
Задание 2	5
Всего	10

10. Образец экзаменационного билета

Теоретические вопросы к экзамену

1. Композиция алгоритмов Взвешенное голосование.
2. Алгоритма AdaBoost. Достоинства и недостатки AdaBoost.
3. Итерации обобщённого алгоритма бустинга AnyBoost.

4. Алгоритм построения композиции классификаторов для голосования по большинству. Преобразование простого голосования во взвешенное.
5. Алгоритм построения композиции для голосования по старшинству. Стратегии выбора параметра алгоритма голосования по старшинству.
6. Типы критериев задачах отбора признаков и моделей. Критерий средней ошибки на контрольных данных, критерий скользящего контроля.
7. Bootstrap.
8. Алгоритм полного перебора всевозможных наборов признаков.
9. Алгоритм Add. Каковы его достоинства и недостатки?
10. Алгоритм Add-Del. Каковы его достоинства и недостатки?
11. Стратегии обхода дерева возможных наборов признаков.
12. Критерии выбора модели, основанные на оценках обобщающей способности.
13. Что строит алгоритм PCA, выполняя обработку всей матрицы наблюдений? Зачем используют PCA? Что такое ковариационная матрица, матрица счетов, матрица нагрузок?
14. Что из себя представляют главные компоненты? Являются ли они коррелированными? Из каких соображений выбирают количество главных компонент, которым стоит ограничиться? Критерия broken stick model.
15. Общая архитектура рекуррентной сети. Особенности архитектуры LSTM-сети, преимущества LSTM-сети.
16. Принцип работы сети Кохонена.
17. Общая архитектура сверточных нейросетей. Двумерная свёртка. Техники свёртки, «фильтр» и «ядро» свёртки. Слои пулинга.
18. Параметрические и непараметрические методы оценивания плотности вероятностей. Смеси.
19. Подход для нахождения оптимальных оценок параметров регрессионной модели. Оценки качества регрессионной модели.
20. Схема однослойной рекуррентной сети. Отличие рекуррентной сети Элмана от предыдущих концепций рекуррентных сетей.
21. Алгоритм k ближайших соседей и его модификация – k взвешенных ближайших соседей. Какие недостатки простейших метрических алгоритмов типа kNN ?
22. Отступ объекта относительно алгоритма. Типы объектов в порядке убывания отступа. Алгоритм STOLP для отбора эталонных объектов.
23. Решающие деревья. Основные этапы алгоритма построения решающего дерева ID3, его преимущества и недостатки. Алгоритм CART. Редукция решающих деревьев.

Образец экзаменационного билета

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»

Физико-технический факультет

<i>Направление подготовки:</i>	09.04.01 Информатика и вычислительная техника
<i>Магистерская программа:</i>	Информатика и вычислительная техника
<i>Программа подготовки:</i>	академическая магистратура
<i>Семестр</i>	3
<i>Учебная дисциплина</i>	Машинное обучение

БИЛЕТ №1

1. Композиция алгоритмов Взвешенное голосование.
2. Общая архитектура рекуррентной сети. Особенности архитектуры LSTM-сети, преимущества LSTM-сети.
3. Рассмотрим задачу классификации с двумя классами 0 и 1. Пусть пространство признаков двумерное. Объекты каждого класса имеют нормальное распределение с математическим ожиданием (0, 0.5) и (0.4, 1) соответственно и матрицей ковариации

$$\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

Априорные вероятности классов равны $P\{Y=0\}=1/3$ и $P\{Y=1\}=2/3$

Найти: уравнение разделяющей поверхности байесова классификатора; собственное разложение матрицы Σ .

Перейти к новым координатам, оси которых совпадают с собственными векторами матрицы Σ . Выписать уравнение разделяющей поверхности байесова классификатора в новых координатах.

Утверждено на заседании кафедры компьютерных технологий,
протокол № ____ от “__” _____ 20__ г.

Зав. кафедрой

Преподаватель

Критерии оценивания экзамена

<i>Номер задания</i>	<i>Количество баллов</i>
Задание 1	10
Задание 2	10
Задание 3	30
Всего	50 баллов

11. Критерии оценивания

По курсу предполагается проведение промежуточной аттестации в виде модульного контроля, выполнение блока лабораторных работ и экзамена. Экзамен сдают студенты с целью повышения рейтинга.

Распределение баллов, которые могут получить студенты в процессе изучения дисциплины

Форма контроля	Максимальное количество баллов
Лабораторная работа №1	20
Лабораторная работа №2	20
Лабораторная работа №3	20
Лабораторная работа №4	20
Модульный контроль	10
Организационно-учебная работа студента	10

Шкала соответствия баллов национальной шкале

Оценка по шкале ECTS	Оценка по 100-балльной шкале	Оценка по государственной шкале (экзамен, дифференцированный зачет)	Оценка по государственной шкале (зачет)
A	90-100	5 (отлично)	зачтено
B	80-89	4 (хорошо)	зачтено
C	75-79	4 (хорошо)	зачтено
D	70-74	3 (удовлетворительно)	зачтено
E	60-69	3 (удовлетворительно)	зачтено
FX	35-59	2 (неудовлетворительно) с возможностью повторной сдачи	не зачтено
F	0-34	2 (неудовлетворительно) с возможностью повторной сдачи при условии обязательного набора дополнительных баллов	не зачтено

12. Материально-техническое обеспечение учебного процесса.

Лекционные занятия проводятся в аудитории, оснащенной мультимедийной техникой и доской. Лабораторные занятия проводятся в компьютерном классе, оборудованном компьютерами с лицензионным программным обеспечением, доступом к сети Интернет, столами, доской.

13. Рекомендованная литература

№ п/п	Наименование	Кол-во экземпляров в библиотеке ДонНУ	Наличие электронной версии в ЭБС
Основная литература			
1.	Ермоленко Т.В. Методы машинного обучения [Электронный ресурс]: учебное пособие / Т.В. Ермоленко – Донецк: ДонНУ, 2017. – Электронные данные (1 файл).	-	+
2.	Ермоленко Т.В. Введение в машинное обучение [Электронный ресурс]: учебное-методическое пособие / Т.В. Ермоленко – Донецк: ДонНУ, 2017. – Электронные данные (1 файл).	-	+
Дополнительная литература			
3.	Коэльо Л.П. Построение систем машинного обучения на языке Python [Электронная книга] / Л.П. Коэльо, В. Ричарт. – М.: Издательство «ДМК Пресс», 2016. – 302 с. URL: http://mirknig.su/knigi/programming/10054-postroenie-sistem-mashinnogo-obucheniya-na-yazyke-python-2-e-izdanie.html (в свободном доступе)	-	-
4.	Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Электронный ресурс]. / В.К. Шитиков, С.Э. Мастицкий – 2017. – 351 с. – Электронные данные, URL: https://github.com/ranalytics/data-mining (в свободном доступе)	-	-

14. Информационные ресурсы

1. К.В. Воронцов. Машинное обучение (курс лекций). URL: <http://www.machinelearning.ru/wiki/index.php?title=ML> (дата обращения 03.01.2017).
2. Портал статей по применению ИТ и машинному обучению - http://habrahabr.ru/hub/machine_learning/ (дата обращения 03.01.2017).
3. Портал искусственного интеллекта URL: <http://www.aiportal.ru/downloads> (дата обращения 03.01.2017).

15. Программное обеспечение

1. Язык R – свободная программная среда для статистических вычислений и графики (open source, лицензия GNU GPL 2).
2. RStudio – интегрированная среда разработки (IDE) для R (open source, лицензия GNU GPL 2).

Рабочая программа рассмотрена и переутверждена на заседании кафедры компьютерных технологий с изменениями (без изменений) на 201 8/19 учебный год.

Протокол № 2 от “30” 08 2018г.

Зав. кафедрой



Т.В. Ермоленко