

Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

НЕЙРОННЫЕ СЕТИ И ИХ ПРИМЕНЕНИЕ

Методические указания к лабораторным работам и организации
самостоятельной работы для студентов направления
«Программная инженерия»
(уровень магистратуры)

2018

**Голубева Александра Александровна,
Аксенов Сергей Владимирович**

Нейронные сети и их применение: Методические указания к лабораторным работам и организации самостоятельной работы для студентов направления «Программная инженерия» (уровень магистратуры) / А.А. Голубева, С.В. Аксенов. – Томск, 2018. – 13 с.

© Томский государственный университет
систем управления и радиоэлектроники, 2018

© Голубева А.А., Аксенов С.В., 2018

Оглавление

1 Введение	4
2 Методические указания к проведению лабораторных работ.....	5
2.1 Лабораторная работа «Поиск наилучшего классификатора для решения задачи бинарной классификации при использовании нескольких типов моделей обучения и сравнительной оценке их результатов».....	5
2.2 Лабораторная работа «Определение оптимального числа кластеров при кластеризации»	7
2.3 Лабораторная работа «Построение модели линейной регрессии»9	
2.4. Лабораторная работа «Уменьшение размерности входного пространства»	11
2.5. Лабораторная работа «Удаление выбросов и определения взаимного влияния признаков»	13
3 Методические указания для организации самостоятельной работы.....	15
3.1 Общие положения	15
3.2 Изучение тем теоретической части дисциплины, вынесенных для самостоятельной проработки	16
3.3 Подготовка к лабораторным работам	16
3.4 Выполнение индивидуального (творческого) задания (ИЗ)	17
3.5 Подготовка реферата.....	18
4 Основная и дополнительная литература.....	19

1 Введение

Методические указания к лабораторным работам и организации самостоятельной работы посвящены реализации учебно-методической поддержке дисциплины «Нейронные сети и их применение».

Данное пособие содержит задания для выполнения лабораторных и самостоятельных работ позволяющие сформировать у магистров практические и теоретические навыки по построению и работе с нейронными сетями.

2 Методические указания к проведению лабораторных работ

2.1 Лабораторная работа «Поиск наилучшего классификатора для решения задачи бинарной классификации при использовании нескольких типов моделей обучения и сравнительной оценке их результатов»

Цель работы

Целью лабораторной работы является приобретение практических и теоретических навыков по поиску наилучшего классификатора для решения задачи бинарной классификации при использовании нескольких типов нечетких моделей.

Рекомендации по подготовке к занятию

Во время подготовки к лабораторной работе рекомендуется ознакомиться с лекционным материалом и просмотреть дополнительную учебно-методическую литературу, рекомендуемую лектором во время проведения занятия.

Содержание занятия / порядок проведения лабораторной работы

1. Получить у преподавателя тестовую выборку данных для решения задачи бинарной классификации.

Структура выборки:

Name	Data Type	Meas.	Description
Sex	nominal		M, F, and I (infant)
Length	continuous	mm	Longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)

Shell weight	continuous	grams	after being dried
Rings	integer		+1.5 gives the age in years

Пример выборки:

M,0.455,0.365,0.095,0.514,0.2245,0.101,0.15,15

M,0.35,0.265,0.09,0.2255,0.0995,0.0485,0.07,7

F,0.53,0.42,0.135,0.677,0.2565,0.1415,0.21,9

M,0.44,0.365,0.125,0.516,0.2155,0.114,0.155,10

I,0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7

I,0.425,0.3,0.095,0.3515,0.141,0.0775,0.12,8

F,0.53,0.415,0.15,0.7775,0.237,0.1415,0.33,20

F,0.545,0.425,0.125,0.768,0.294,0.1495,0.26,16

M,0.475,0.37,0.125,0.5095,0.2165,0.1125,0.165,9

F,0.55,0.44,0.15,0.8945,0.3145,0.151,0.32,19

F,0.525,0.38,0.14,0.6065,0.194,0.1475,0.21,14

M,0.43,0.35,0.11,0.406,0.1675,0.081,0.135,10

M,0.49,0.38,0.135,0.5415,0.2175,0.095,0.19,11

F,0.535,0.405,0.145,0.6845,0.2725,0.171,0.205,10

F,0.47,0.355,0.1,0.4755,0.1675,0.0805,0.185,10

2. Отобрать из выборки примеры, принадлежащие двум классам и использовать только отобранные примеры для построения классификатора. В качестве признаков берутся только численные атрибуты.

3. Разделить выборку на две части: обучающую (80% образцов) и тестирующую (20% образцов).

4. Создать классификаторы для предметной области на основе трех моделей (наивный Байес, многослойный перцептрон, и деревья решений) с использованием обучающей выборки.

5. Оценить качество работы классификаторов на тестирующей выборке – получить для каждого класса значения полноты, точности и точности для всей модели.

6. Выбрать лучший классификатор.

7. Подготовить отчет о проделанной работе.

Содержание отчета

1. Титульный лист.
2. Описание задания.
3. Описание исследуемого набора данных (краткая характеристика набора, включающая описание атрибутов, использующихся при анализе).
4. Описание обученных классификаторов. Для деревьев решений: правила или вид дерева, для нейронной сети: характеристики нейронов, включая значения их весов, для наивного Байеса: вид или характеристики функций распределений.
5. Описание результатов оценки качества классификаторов.
6. Заключение, включающее выбор наилучшего классификатора.

2.2 Лабораторная работа «Определение оптимального числа кластеров при кластеризации»

Цель работы

Целью лабораторной работы является приобретение практических и теоретических навыков по определению оптимального числа кластеров при кластеризации данных.

Рекомендации по подготовке к занятию

Во время подготовки к лабораторной работе рекомендуется ознакомиться с лекционным материалом и просмотреть дополнительную учебно-методическую литературу, рекомендуемую лектором во время проведения лабораторной работы.

Содержание занятия / порядок проведения лабораторной работы

1. Получить у преподавателя тестовую выборку данных для решения задачи определения оптимального числа кластеров, содержащую как минимум 300 объектов в выборке и количество числовых атрибутов, которые подлежат анализу не менее 6.

Структура выборки:

Name	Data Type	Meas.	Description
Sex	nominal		M, F, and I (infant)

Length	continuous	mm	Longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried
Rings	integer		+1.5 gives the age in years

Пример выборки:

M,0.455,0.365,0.095,0.514,0.2245,0.101,0.15,15

M,0.35,0.265,0.09,0.2255,0.0995,0.0485,0.07,7

F,0.53,0.42,0.135,0.677,0.2565,0.1415,0.21,9

M,0.44,0.365,0.125,0.516,0.2155,0.114,0.155,10

I,0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7

I,0.425,0.3,0.095,0.3515,0.141,0.0775,0.12,8

F,0.53,0.415,0.15,0.7775,0.237,0.1415,0.33,20

F,0.545,0.425,0.125,0.768,0.294,0.1495,0.26,16

M,0.475,0.37,0.125,0.5095,0.2165,0.1125,0.165,9

F,0.55,0.44,0.15,0.8945,0.3145,0.151,0.32,19

F,0.525,0.38,0.14,0.6065,0.194,0.1475,0.21,14

M,0.43,0.35,0.11,0.406,0.1675,0.081,0.135,10

M,0.49,0.38,0.135,0.5415,0.2175,0.095,0.19,11

F,0.535,0.405,0.145,0.6845,0.2725,0.171,0.205,10

F,0.47,0.355,0.1,0.4755,0.1675,0.0805,0.185,10

2. Построить процесс кластеризации набора выбранных числовых атрибутов с помощью алгоритма K-средних.

3. Провести исследование загруженного набора данных на предмет нахождения оптимального числа кластеров с учетом метода оценки качества. Предполагается, что данные разбиваются на N кластеров, при этом N меняется от 2 до 12. Исходя из значений компактности, выбрать оптимальное значение k с помощью метода локтя.

4. Подготовить отчет о проделанной работе.

Содержание отчета

1. Титульный лист.

2. Описание задания.

3. Описание исследуемого набора данных (краткая характеристика набора, включающая описание атрибутов, используемых при анализе).

4. Описание результатов эксперимента с оценкой компактности для разного количества кластеров.

5. Выводы по кластеризации исследуемого набора. Привести результат (количество кластеров и его координаты) кластеризации с наилучшим значением критерия качества (компактности).

2.3 Лабораторная работа «Построение модели линейной регрессии»

Цель работы

Целью лабораторной работы является приобретение практических и теоретических навыков по построению модели линейной регрессии.

Рекомендации по подготовке к занятию

Во время подготовки к лабораторной работе рекомендуется ознакомиться с лекционным материалом и просмотреть дополнительную учебно-методическую литературу, рекомендуемую лектором во время проведения лабораторной работы.

Содержание занятия / порядок проведения лабораторной работы

1. Получить у преподавателя тестовую выборку данных для решения задачи построения модели линейной регрессии, который содержит

как минимум 150 объектов в выборке и количество числовых атрибутов, которые подлежат анализу не менее 4.

Структура выборки:

Name	Data Type	Meas.	Description
Sex	nominal		M, F, and I (infant)
Length	continuous	mm	Longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried
Rings	integer		+1.5 gives the age in years

Пример выборки:

M,0.455,0.365,0.095,0.514,0.2245,0.101,0.15,15
M,0.35,0.265,0.09,0.2255,0.0995,0.0485,0.07,7
F,0.53,0.42,0.135,0.677,0.2565,0.1415,0.21,9
M,0.44,0.365,0.125,0.516,0.2155,0.114,0.155,10
I,0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7
I,0.425,0.3,0.095,0.3515,0.141,0.0775,0.12,8
F,0.53,0.415,0.15,0.7775,0.237,0.1415,0.33,20
F,0.545,0.425,0.125,0.768,0.294,0.1495,0.26,16
M,0.475,0.37,0.125,0.5095,0.2165,0.1125,0.165,9
F,0.55,0.44,0.15,0.8945,0.3145,0.151,0.32,19
F,0.525,0.38,0.14,0.6065,0.194,0.1475,0.21,14
M,0.43,0.35,0.11,0.406,0.1675,0.081,0.135,10
M,0.49,0.38,0.135,0.5415,0.2175,0.095,0.19,11
F,0.535,0.405,0.145,0.6845,0.2725,0.171,0.205,10
F,0.47,0.355,0.1,0.4755,0.1675,0.0805,0.185,10

2. Построить в процесс работы модели линейной регрессии (выборку разделить на две части 70% - обучение, 30% тест).
3. Получить уравнение регрессии, статистические характеристики параметров модели, а также значения среднеквадратичной ошибки и коэффициента линейной регрессии для тестовой выборки.
4. Подготовить отчет о проделанной работе.

Содержание отчета

1. Титульный лист.
2. Описание задания.
3. Описание исследуемого набора данных (краткая характеристика набора, включающая описание атрибутов, используемых при анализе).
4. Описание результатов моделирования.

2.4. Лабораторная работа «Уменьшение размерности входного пространства»

Цель работы

Целью лабораторной работы является приобретение практических и теоретических навыков по уменьшению размерности входного пространства.

Рекомендации по подготовке к занятию

Во время подготовки к лабораторной работе рекомендуется ознакомиться с лекционным материалом и просмотреть дополнительную учебно-методическую литературу, рекомендуемую лектором во время проведения лабораторной работы.

Содержание занятия / порядок проведения лабораторной работы

1. Получить у преподавателя тестовую выборку данных для решения задачи (регрессии, кластеризации или классификации), который со-

держит как минимум 350 объектов в выборке и количество числовых атрибутов, которые подлежат анализу не менее 12.

Структура выборки:

Name	Data Type	Meas.	Description
Sex	nominal		M, F, and I (infant)
Length	continuous	mm	Longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried
Rings	integer		+1.5 gives the age in years

Пример выборки:

M,0.455,0.365,0.095,0.514,0.2245,0.101,0.15,15
M,0.35,0.265,0.09,0.2255,0.0995,0.0485,0.07,7
F,0.53,0.42,0.135,0.677,0.2565,0.1415,0.21,9
M,0.44,0.365,0.125,0.516,0.2155,0.114,0.155,10
I,0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7
I,0.425,0.3,0.095,0.3515,0.141,0.0775,0.12,8
F,0.53,0.415,0.15,0.7775,0.237,0.1415,0.33,20
F,0.545,0.425,0.125,0.768,0.294,0.1495,0.26,16
M,0.475,0.37,0.125,0.5095,0.2165,0.1125,0.165,9
F,0.55,0.44,0.15,0.8945,0.3145,0.151,0.32,19
F,0.525,0.38,0.14,0.6065,0.194,0.1475,0.21,14
M,0.43,0.35,0.11,0.406,0.1675,0.081,0.135,10
M,0.49,0.38,0.135,0.5415,0.2175,0.095,0.19,11
F,0.535,0.405,0.145,0.6845,0.2725,0.171,0.205,10
F,0.47,0.355,0.1,0.4755,0.1675,0.0805,0.185,10

2. Построить процесс работы модели выделения главных компонент PCA и SOM.

3. Для PCA получить значения выборочной дисперсии, процента дисперсии компонент, кумулятивного процента дисперсии, формулы для получения главных компонент, 20 первых векторов новых признаков главных компонент для исследуемой выборки. Для SOM разброс по кластерам, а также 20 первых векторов новых признаков.

4. Подготовить отчет о проделанной работе.

Содержание отчета

1. Титульный лист.
2. Описание задания.
3. Описание исследуемого набора данных (краткая характеристика набора, включающая описание атрибутов, используемых при анализе).
4. Описание результатов моделирования.

2.5. Лабораторная работа «Удаление выбросов и определения взаимного влияния признаков»

Цель работы

Целью лабораторной работы является приобретение практических и теоретических навыков по удалению выбросов и определению взаимного влияния признаков.

Рекомендации по подготовке к занятию

Во время подготовки к лабораторной работе рекомендуется ознакомиться с лекционным материалом и просмотреть дополнительную учебно-методическую литературу, рекомендуемую лектором во время проведения лабораторной работы.

Содержание занятия / порядок проведения лабораторной работы

1. Получить у преподавателя тестовую выборку данных для решения задачи (регрессии, кластеризации или классификации), который со-

держит как минимум 350 объектов в выборке и количество числовых атрибутов, которые подлежат анализу не менее 5.

Структура выборки:

Name	Data Type	Meas.	Description
Sex	nominal		M, F, and I (infant)
Length	continuous	mm	Longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried
Rings	integer		+1.5 gives the age in years

Пример выборки:

M,0.455,0.365,0.095,0.514,0.2245,0.101,0.15,15
M,0.35,0.265,0.09,0.2255,0.0995,0.0485,0.07,7
F,0.53,0.42,0.135,0.677,0.2565,0.1415,0.21,9
M,0.44,0.365,0.125,0.516,0.2155,0.114,0.155,10
I,0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7
I,0.425,0.3,0.095,0.3515,0.141,0.0775,0.12,8
F,0.53,0.415,0.15,0.7775,0.237,0.1415,0.33,20
F,0.545,0.425,0.125,0.768,0.294,0.1495,0.26,16
M,0.475,0.37,0.125,0.5095,0.2165,0.1125,0.165,9
F,0.55,0.44,0.15,0.8945,0.3145,0.151,0.32,19
F,0.525,0.38,0.14,0.6065,0.194,0.1475,0.21,14
M,0.43,0.35,0.11,0.406,0.1675,0.081,0.135,10
M,0.49,0.38,0.135,0.5415,0.2175,0.095,0.19,11
F,0.535,0.405,0.145,0.6845,0.2725,0.171,0.205,10

F,0.47,0.355,0.1,0.4755,0.1675,0.0805,0.185,10

2. Построить процесс работы модели построения корреляционной матрицы и определения выбросов.

3. По корреляционной матрице определить признаки, которые зависят друг от друга линейно (значение в матрице по модулю больше 0.85) и определить вид зависимости: отрицательная или положительная. Для решения задачи определения выбросов установить количество выбросов равным 5% и путем изменения количества соседей $K=3,4,5$ получить номера сущностей, которые являются выбросами. Выбрать признаки, которые обозначены как выбросы три раза, два раза и один раз при разных K .

4. Подготовить отчет о проделанной работе.

Содержание отчета

1. Титульный лист.
2. Описание задания.
3. Описание исследуемого набора данных (краткая характеристика набора, включающая описание атрибутов, использующихся при анализе).
4. Описание результатов моделирования.

3 Методические указания для организации самостоятельной работы

3.1 Общие положения

Целями самостоятельной работы являются систематизация, расширение и закрепление теоретических знаний.

Самостоятельная работа студента по дисциплине «Параллельные вычисления и системы» включает следующие виды активности:

1. Изучение тем теоретической части дисциплины, вынесенных для самостоятельной проработки.
2. Подготовка к лабораторным работам.
3. Подготовка реферата.
4. Выполнение индивидуальных заданий.

3.2 Изучение тем теоретической части дисциплины, вынесенных для самостоятельной проработки

Изучение тем теоретической части дисциплины, вынесенных для самостоятельной проработки:

- Ядерные методы. В рамках данного раздела рекомендуется обратить внимание на: метод ядерных оценок, на вероятностные нейронные сети, обобщенно-регрессионные нейронные сети, методы ядерной аппроксимации.

- Вероятностные модели для категориальных данных.

В рамках данного раздела рекомендуется обратить внимание на структуру модели, особенности и характеристики.

- Дискриминантное обучение путем оптимизации условного правдоподобия. В рамках данного раздела рекомендуется обратить внимание на дискриминантную модель логической регрессии.

- Вероятностные модели со скрытыми переменными. В рамках данного раздела рекомендуется обратить внимание на решение четырехкаскадной задачи классификации.

Рекомендуемая литература:

1. Яхьяева Г.Э. Нечеткие множества и нейронные сети: учеб. пособие. – 2-е изд., испр. - М.: БИНОМ; Лаборатория знаний, 2008. – 315 с. В библиотеке ТУСУРа: 1 экз.

2. Рутковская Д.А. Нейронные сети, генетические алгоритмы и нечеткие системы: пер. с польск. / Д. Рутковская, М. Пилиньский, Л. Рутковский пер. И.Д. Рудинский. – М.: Горячая линия-Телеком, 2006. – 383 с. В библиотеке ТУСУРа: 20 экз.

3.3 Подготовка к лабораторным работам

В рамках выполнения подготовки к лабораторным работам рекомендуется детально познакомиться с теоретическим материалом по темам лабораторных работ.

Рекомендуемая литература:

1. Яхьяева Г.Э. Нечеткие множества и нейронные сети: учеб. пособие. – 2-е изд., испр. - М.: БИНОМ; Лаборатория знаний, 2008. – 315 с. В библиотеке ТУСУРа: 1 экз.

2. Рутковская Д.А. Нейронные сети, генетические алгоритмы и нечеткие системы: пер. с польск. / Д. Рутковская, М. Пилиньский, Л. Рутковский пер. И.Д. Рудинский. – М.: Горячая линия-Телеком, 2006. – 383 с. В библиотеке ТУСУРа: 20 экз.

3.4 Выполнение индивидуального (творческого) задания (ИЗ)

В рамках выполнения индивидуального (творческого) задания (ИЗ) необходимо подготовить 7 минутный доклад, раскрывающий одну из следующих тем:

Вариант индивидуального задания определяется преподавателем в индивидуальном порядке, основываясь на уровень знаний и студента.

Примеры тем:

1. Иерархическая кластеризация
2. Метод опорных векторов (SVM –support vector machines)
3. Линейный дискриминантный анализ (LDA - Linear discriminant analysis)
4. Алгоритм T-SNe (T-distributed stochastic neighbor embedding)
5. EM-алгоритм (Expectation-maximization algorithm)
6. Алгоритм AdaBoost
7. CART: Classification and Regression Trees
8. Скрытые Марковские модели (Hidden Markov Models)
9. Сингулярное разложение матрицы (SVD Singular-value decomposition)
10. Алгоритмы кластеризации на основе сеток и плотности (Grid-based and density based clustering)
11. Алгоритмы кластеризации данных большой размерности (CLIQUE, PROCLUS, ORCLUS)
12. Анализ выбросов (Outlier analysis)
13. Методы обучения с частичным привлечением учителя (semi-supervised learning)

14. Методы кластеризации текстовых данных (Clustering methods for texts)
15. Алгоритмы кластеризация временных рядов (Time series clustering)
16. Методы классификации временных рядов (Time series classification)
17. Кластеризация графов (Graph clustering)
18. Алгоритм поиска ассоциативных правил FP-growth (Frequent Pattern-Growth)

3.5 Подготовка реферата

Подготовка реферата по одной из тем:

1. Использование модификаций генетического алгоритма при решении задачи глобальной оптимизации.
 2. Алгоритмы настройки нейро-нечетких систем вывода
 3. Модель иерархической временной памяти НТМ.
 4. Осцилляторные нейронные сети.
 5. Рекуррентные нейронные сети.
 6. Кластеризация графов.
 7. Применение нейронных сетей для решения задачи сегментации визуальных сцен.
 8. Восстановление трехмерных образов с помощью нейронных сетей.
 9. Сети адаптивного резонанса
 10. Процедуры настройки RBF-сетей.
- Рекомендуемый объем реферата не более 20 страниц.

4 Основная и дополнительная литература

Основная литература:

1. Аксенов С.В. Организация и использование нейронных сетей (методы и технологии): науч. издание / С.В. Аксенов, В.Б. Новосельцев; ред. В.Б. Новосельцев. – Томск: Изд-во НТЛ, 2006. – 126 с. В библиотеке ТУСУРа: 15 экз.

2. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилинский, Л. Рутковский; пер. с польск. И.Д. Рудинского. – 2-е изд., стереотип. – М.: Горячая линия – Телеком, 2013. – 384 с. [Электронный ресурс]: ЭБС ЛАНЬ. – URL: <http://e.lanbook.com/view/book/11843/>

3. Замятин Н.В. Нечеткая логика и нейронные сети: учеб. пособие. – Томск: Эль Контент, 2014. – 146 с. В библиотеке ТУСУРа: 10 экз.

Дополнительная литература:

1. Усков А.А. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика / А.А. Усков, А.В. Кузьмин. – М.: Горячая линия-Телеком, 2004. – 143 с. В библиотеке ТУСУРа: 50 экз.

2. Рутковская Д.А. Нейронные сети, генетические алгоритмы и нечеткие системы: пер. с польск. / Д. Рутковская, М. Пилинский, Л. Рутковский пер. И.Д. Рудинский. – М.: Горячая линия-Телеком, 2006. – 383 с. В библиотеке ТУСУРа: 20 экз.