**Лабораторная работа 2**

**по дисциплине «Нейросетевые технологии»**

**3 курс 6 семестр**

**Классификация на 2 класса по косвенным признакам**

Целью данной работы является демонстрация способности нейронных сетей (НС) решать неформализованные задачи.

**Общие положения. Выбор вариантов**

Данная работа выполняется на нейроимитаторе с помощью программы Neural Network Wizard на ПК.

Каждый студент должен получить индивидуальный вариант задания. Выбрав вариант, необходимо выделить входные данные, выходные данные и определить алгоритм создания обучающей выборки. Отчет о проделанной самостоятельной работе выполняется в печатном виде.

На титульном листе отчета следует указать:

* наименование учебной дисциплины;
* номер и название лабораторной работы;
* номер варианта;
* курс и номер учебной группы;
* фамилию, имя, отчество автора;
* фамилию, имя, отчество преподавателя.

Выбор варианта задания выполняется по согласованию с преподавателем.

**Указания по выполнению задания**

В начале работы следует привести полную формулировку задания (как его общую часть, так и текст индивидуального варианта).

Обучающую выборку рекомендуется формировать в программе Microsoft Excel. Этап выбора топологии НС, проведения обучения НС и оценки точности её работы (допустима относительная погрешность выхода НС 5-10%) на «зашумлённых» входных данных проводится с помощью программы Neural Network Wizard. При обучении нейросети на нейроимитаторе рекомендуется опробовать несколько вариантов обучения:

1. С автоматической нормализацией входных и целевых полей либо с нормализацией по формулам.
2. С использованием критериев остановки обучения либо с остановкой обучения вручную.

Высоко будет оцениваться субъективное мнение студента в части оценки эффективности проведенного обучения нейросети.

**Задание**

Сеть необходимо обучить классификации на 2 и более класса по набору косвенных признаков.

* + Необходимо выбрать задачу и согласовать её с преподавателем. Примерами таких задач могут служить следующие: «Мужчина – женщина», «Студент – преподаватель», «Студент живущий дома – в общежитии», «Семейный – холостой», «Экстраверт – интроверт» и т.д.
  + Составить вопросник из 15-20 косвенных вопросов, по ответам на которые, с точки зрения студента, возможно провести классификацию. Список вопросов утверждается преподавателем. Примером косвенного вопроса в задаче «Мужчина – женщина» может служить вопрос «Носите ли Вы дома халат?», однако вопросы «Носите ли Вы дома юбку?» или «Приходится ли Вам по утрам бриться?» косвенными считаться не могут. В задаче «Студент – преподаватель» вопрос «Есть ли у Вас студенческий билет?» является прямым, «Готовитесь ли Вы к лекциям?» – косвенным. Каждый вопрос должен предполагать **три варианта ответа** – «да», «нет», «иногда» (или «не знаю»).
  + Необходимо проанкетировать не менее 20-ти человек по составленному вопроснику.
  + Составить обучающую и тестирующую выборки, зашифровав вопросы, варианты ответов и результаты анкетирования с помощью чисел и буквенных обозначений.
  + Провести пробное обучение НС. Выявить значимые и незначимые вопросы для решения задачи.
  + Провести минимизацию вопросника.
  + Выявить оптимальную топологию НС и провести обучение НС.
  + Окончательный и первоначальный варианты НС продемонстрировать преподавателю.

**Пример выполнения задания**

1 Постановка задачи

Требуется составить тестовый вопросник, позволяющий отнести человека к одному из следующих классов – мужчина или женщина.

2 Составление вопросника

Составляем перечень вопросов, каждый вопрос предполагает три варианта ответа – «да», «нет», «не знаю» («иногда»).

Вопросы:

1. Вы пользуетесь декоративной косметикой?
2. Вы посещаете салоны красоты?
3. Вы курите?
4. Вы любите детей?
5. Вам нравится смотреть мелодраматические сериалы?
6. Вам дарили цветы?
7. Вам приходилось участвовать в драке с мужчиной?
8. Вы отслеживаете график проведения спортивных соревнований и их результаты?
9. Вы носите дома халат?
10. Ваши дети носят (или будут носить) фамилию, полученную Вами при рождении?
11. Вы умеете шить?
12. Вы самостоятельно собираете мебель по имеющимся инструкциям?

3 Формирование обучающей выборки

Далее составляем обучающую выборку, присваивая при этом каждому вопросу, ответу и отвечающему человеку соответствующее числовое значение. Обучающая выборка представлена в таблице 1. Столбцы *q1*–*q12* являются входными параметрами и содержат ответы на вопросы с соответствующими номерами. Столбцы *man* и *wom* – целевые параметры, соответствуют классу мужчин и женщин. Строки – варианты ответов опрашиваемых людей. Числа на пересечении строк и столбцов входных параметров соответствуют данному ответу: 1 – «да», 9 – «нет», 5 – «иногда» («затрудняюсь ответить»). Числа в целевых столбцах – полученный результат в вариативной форме, от 0 до 100%.

Сохраняем выборку в txt-файл с разделителями табуляции.

Таблица 1 – Обучающая выборка

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **q1** | **q2** | **q3** | **q4** | **q5** | **q6** | **q7** | **q8** | **q9** | **q10** | **q11** | **q12** | **man** | **wom** |
| **1** | **5** | **9** | **1** | **5** | **1** | **9** | **9** | **1** | **9** | **1** | **9** | **0** | **100** |
| **1** | **1** | **9** | **1** | **1** | **1** | **9** | **5** | **1** | **9** | **1** | **9** | **0** | **100** |
| **9** | **9** | **9** | **5** | **9** | **9** | **1** | **9** | **9** | **1** | **9** | **1** | **100** | **0** |
| **9** | **5** | **9** | **5** | **9** | **5** | **1** | **1** | **5** | **1** | **9** | **5** | **100** | **0** |
| **1** | **5** | **9** | **5** | **1** | **1** | **9** | **9** | **1** | **5** | **5** | **9** | **0** | **100** |
| **9** | **9** | **1** | **1** | **5** | **9** | **1** | **1** | **9** | **1** | **9** | **1** | **100** | **0** |
| **1** | **1** | **9** | **9** | **5** | **1** | **9** | **9** | **1** | **5** | **9** | **9** | **0** | **100** |
| **5** | **9** | **1** | **1** | **1** | **1** | **5** | **9** | **1** | **9** | **5** | **5** | **0** | **100** |
| **1** | **5** | **1** | **1** | **1** | **1** | **1** | **9** | **1** | **1** | **9** | **9** | **0** | **100** |
| **5** | **5** | **9** | **5** | **9** | **9** | **5** | **5** | **5** | **1** | **9** | **5** | **100** | **0** |
| **9** | **9** | **5** | **5** | **9** | **9** | **1** | **5** | **9** | **1** | **9** | **9** | **100** | **0** |

3 Построение и обучение нейросети на нейроиммитаторе

1. Запускаем нейроиммитатор NNW.
2. В качестве файла с обучающей выборкой указываем txt-файл, сформированный ранее. Нажимаем «Далее».
3. Указываем, что поля *q1*–*q12* являются входными, а *man* и *wom* – целевыми. Остальные параметры – по умолчанию. Нажимаем «Далее».
4. Задаем параметры нейросети. Устанавливаем крутизну передаточной функции нейронов (параметр сигмоиды) – 2. Остальные параметры – по умолчанию (число скрытых слоев – 1, число нейронов в скрытом слое – 10). Нажимаем «Далее».
5. В данном окне при необходимости задаем параметры обучения. В нашем случае оставляем все параметры по умолчанию. Нажимаем «Далее».
6. В этом окне можно просмотреть все заданные параметры. Нажимаем «Далее».
7. В окне обучения системы нажимаем «Пуск обучения». Ждем окончания процесса обучения (рисунок 1) или, при достижении необходимой точности (ошибки), нажимаем «Остановка обучения». Нажимаем «Далее».

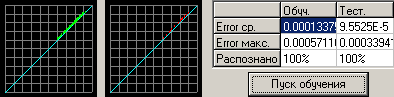


Рисунок 1 – Окончание процесса обучения

1. В окне проверки результатов можно произвести расчет для каких-либо значений переменных *x* и *y*. Для сохранения нейросети нажимаем «Сохранить».

4 Оценка ошибки работы нейросети

В качестве подтверждающей выборки используется сама обучаюая выборка. Результаты работы нейросети с подтверждающей выборкой представлены в таблице 2. Ошибка работы оценивается по количеству верных совпадений.

Согласно данным таблицы 2 относительная погрешность работы нейросети с подтверждающей выборкой составляет не более 0,65%.

Для формирования тестовой выборки проводим опрос среди студентов, их ответы предлагаем обученной сети в качестве тестовых.

Результаты работы нейросети с тестовыми примерами представлены в таблице 3.

Таблица 2 – Результат обучения сети на подтверждающей выборке

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **q1** | **q2** | **q3** | **q4** | **q5** | **q6** | **q7** | **q8** | **q9** | **q10** | **q11** | **q12** | **man** | **wom** |
| **1** | **5** | **9** | **1** | **5** | **1** | **9** | **9** | **1** | **9** | **1** | **9** | **4.9** | **99.9** |
| **1** | **1** | **9** | **1** | **1** | **1** | **9** | **5** | **1** | **9** | **1** | **9** | **-2.9** | **99.9** |
| **9** | **9** | **9** | **5** | **9** | **9** | **1** | **9** | **9** | **1** | **9** | **1** | **85.7** | **13.8** |
| **9** | **5** | **9** | **5** | **9** | **5** | **1** | **1** | **5** | **1** | **9** | **5** | **99.9** | **-1.6** |
| **1** | **5** | **9** | **5** | **1** | **1** | **9** | **9** | **1** | **5** | **5** | **9** | **5.2** | **100** |
| **9** | **9** | **1** | **1** | **5** | **9** | **1** | **1** | **9** | **1** | **9** | **1** | **100** | **8.26** |
| **1** | **1** | **9** | **9** | **5** | **1** | **9** | **9** | **1** | **5** | **9** | **9** | **-2.2** | **99.9** |
| **5** | **9** | **1** | **1** | **1** | **1** | **5** | **9** | **1** | **9** | **5** | **5** | **17** | **88.2** |
| **1** | **5** | **1** | **1** | **1** | **1** | **1** | **9** | **1** | **1** | **9** | **9** | **-9.1** | **99.9** |
| **5** | **5** | **9** | **5** | **9** | **9** | **5** | **5** | **5** | **1** | **9** | **5** | **77.3** | **15.8** |
| **9** | **9** | **5** | **5** | **9** | **9** | **1** | **5** | **9** | **1** | **9** | **9** | **100** | **1.3** |

Из таблицы 3 видно, что сеть обучилась достаточно хорошо, совпадения по всем пунктам. Анализируя полученные результаты, можно сделать вывод, что вопросы 4 и 11, по-видимому, являются малоинформативными. Ответы на вопрос 4 содержат значительное количество неопределенных ответов, притом, что остальные значения в этом поле в подавляющем большинстве одинаковы («да»). 11-й вопрос содержит наибольшее количество одинаковых ответов («нет»). Следовательно, их удаление из обучающей выборки и списка вопросов не должно привести к существенному искажению в обучении и работе сети. Проверим это предположение экспериментально. После корректировки обучающей выборки проводим новое обучение нейронной сети. Результаты показаны в таблице 4. Результаты распознавания не ухудшились.

Таблица 3 – Результат повторного обучения сети на тестовой выборке

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **q1** | **q2** | **q3** | **q5** | **q6** | **q7** | **q8** | **q9** | **q10** | **q12** | **man** | **wom** |
| **1** | **5** | **9** | **5** | **1** | **9** | **9** | **1** | **9** | **9** | **4.9** | **99.9** |
| **1** | **1** | **9** | **1** | **1** | **9** | **5** | **1** | **9** | **9** | **-2.9** | **99.9** |
| **9** | **9** | **9** | **9** | **9** | **1** | **9** | **9** | **1** | **1** | **85.7** | **13.8** |
| **9** | **5** | **9** | **9** | **5** | **1** | **1** | **5** | **1** | **5** | **99.9** | **-1.6** |
| **1** | **5** | **9** | **1** | **1** | **9** | **9** | **1** | **5** | **9** | **5.2** | **100** |
| **9** | **9** | **1** | **5** | **9** | **1** | **1** | **9** | **1** | **1** | **100** | **8.26** |
| **1** | **1** | **9** | **5** | **1** | **9** | **9** | **1** | **5** | **9** | **-2.2** | **99.9** |
| **5** | **9** | **1** | **1** | **1** | **5** | **9** | **1** | **9** | **5** | **17** | **88.2** |
| **1** | **5** | **1** | **1** | **1** | **1** | **9** | **1** | **1** | **9** | **-9.1** | **99.9** |
| **5** | **5** | **9** | **9** | **9** | **5** | **5** | **5** | **1** | **5** | **77.3** | **15.8** |
| **9** | **9** | **5** | **9** | **9** | **1** | **5** | **9** | **1** | **9** | **100** | **1.3** |

Оценку работы обученной сети проведем путем тестирования реальных людей – мужчин и женщин. Результаты представлены в таблице 4. Первые четыре строки ответов – женские, последующие – мужские.

Таблица 4 – Результат обучения сети на тестовых примерах

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **q1** | **q2** | **q3** | **q5** | **q6** | **q7** | **q8** | **q9** | **q10** | **q12** | **man** | **wom** |
| **9** | **9** | **1** | **9** | **1** | **1** | **1** | **5** | **9** | **9** | **63.4** | **61.4** |
| **9** | **1** | **9** | **1** | **1** | **9** | **5** | **1** | **9** | **9** | **25.92** | **96.31** |
| **1** | **5** | **9** | **1** | **1** | **5** | **9** | **1** | **9** | **9** | **0** | **100.2** |
| **1** | **5** | **9** | **1** | **1** | **9** | **9** | **5** | **9** | **9** | **14.5** | **97.7** |
| **9** | **9** | **1** | **9** | **5** | **1** | **5** | **9** | **1** | **1** | **99.5** | **0.33** |
| **5** | **9** | **1** | **5** | **5** | **5** | **5** | **9** | **1** | **1** | **64.6** | **27.3** |
| **9** | **9** | **9** | **9** | **9** | **9** | **9** | **9** | **5** | **5** | **80.2** | **42.9** |
| **9** | **9** | **5** | **5** | **5** | **5** | **1** | **5** | **1** | **5** | **83.9** | **20.4** |

По данным таблицы 4 видно, что во всех случаях, кроме первого, нейросеть однозначно и правильно указывала категорию, к которой должен относиться отвечающий человек. Кроме того, по вероятности, с которой машина относит того или иного человека к категории, можно косвенно судить о наличии в характере мужских и женских черт.

Обучение проведено успешно, обучающая выборка составлена корректно.