**Лабораторная работа 3**

**Классификация с помощью нейронных сетей**

*Нейронные сети* (НС) представляют собой вычислительные структуры, моделирующие простые биологические процессы, аналогичные процессам, происходящим в человеческом мозге. Нейросети – это распределенные и параллельные системы, способные к адаптивному обучению путем реакции на положительные и отрицательные воздействия. В основе построения сети лежит элементарный преобразователь, называемый *искусственным нейроном* или просто *нейроном* по аналогии с его биологическим прототипом. Структуру нейросети можно описать следующим образом. Нейросеть состоит из нескольких слоев: входной, внутренние (скрытые) и выходной слои. Входной слой реализует связь с входными данными, выходной – с выходными. Внутренних слоев может быть от одного и больше. В каждом слое содержится несколько нейронов.

Рассмотрим классификацию на классическом примере *ирисов Фишера* (150 цветов трех сортов; каждый экземпляр описывается четырьмя параметрами: длина и ширина чашелистика, длина и ширина лепестка).

1. При обращении к Deductor Academic появляется окно (рис.1).



Рис.1 Исходное окно

После нажатия ОК появляется первое окно, предоставляющее очень мало возможностей. С помощью мастера импорта загрузить в *Deductor* текстовый файл с исходными данными. В приложении приведены исходные данные, которые нужно перевести в текстовый файл.

2.Мастер импорта в этой задаче содержит 10 шагов. Все шаги с этим мастером аналогичны шагам, выполненным в ПЗ\_2.

После нажатия «Готово» на 10-ом шаге появляется результат работы мастера импорта (рис.2).

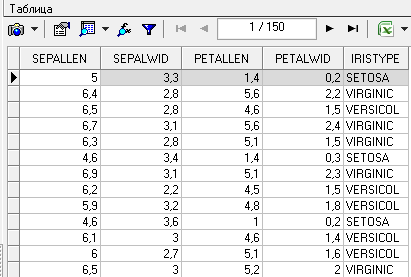


Рис.2 Фрагмент таблицы данных

3.Мастер обработки вводится нажатием на выделенный блок в строке мастеров. Здесь мастер обработки имеет 9 шагов.

Выделяется опция «Нейросеть» в окне Data Mining (рис.3).

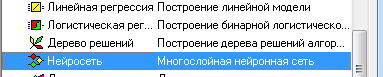


Рис.3 Выделение нейросети

На 2-ом шаге (важный шаг, поэтому нужно выполнить его внимательно) вначале настраиваются столбцы (рис.4).

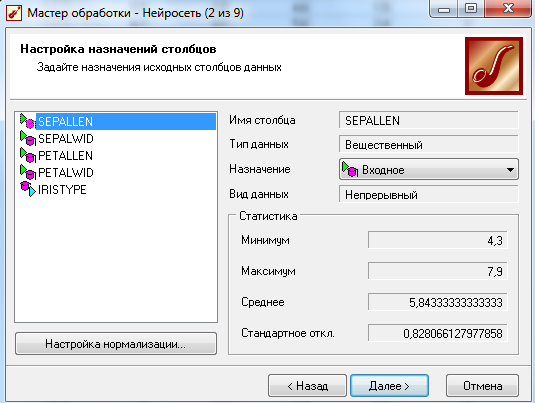


Рис.4 Настройка столбцов

Все входные переменные настраиваются одинаково. Выходной - иначе (рис.5). Вначале выделить тип цветка *IRISTYPE*.

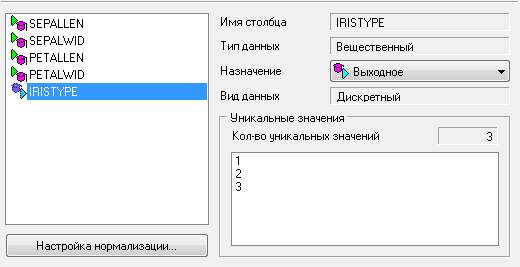


Рис.5 Настройка выходного параметра

Далее на этом же шаге нажать «Настройка нормализации» и выделить выходной параметр, в котором установить опцию «Уникальные значения» (рис.6).

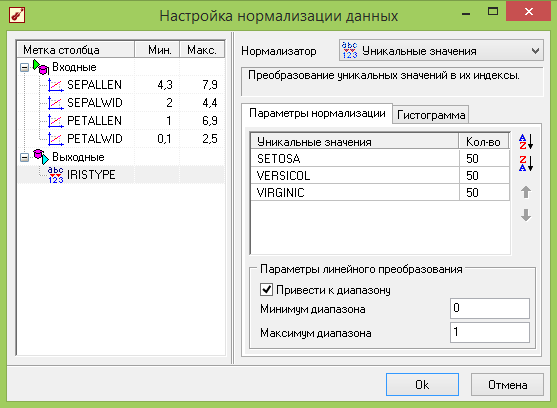


Рис.6 Настройка нормализации для выходного параметра

На 3-ем шаге разбить все данные на обучающее и тестовое множества в отношении 70% и 30% (105 и 45 наблюдений, соответственно, рис.7).

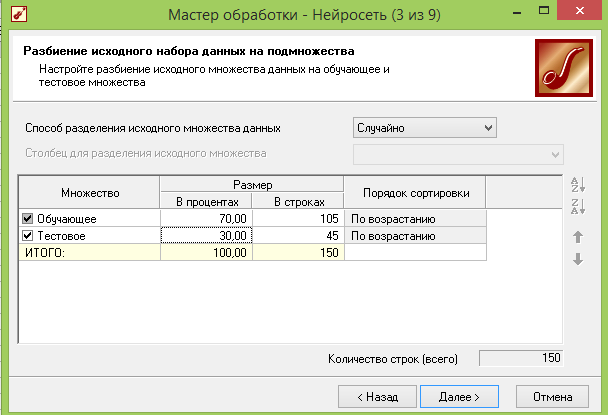


Рис.7 Обучающее и тестовое множества

На 4-ом шаге появляется структура нейронной сети, в которой увеличим до 4 число нейронов в скрытом слое (рис.8).

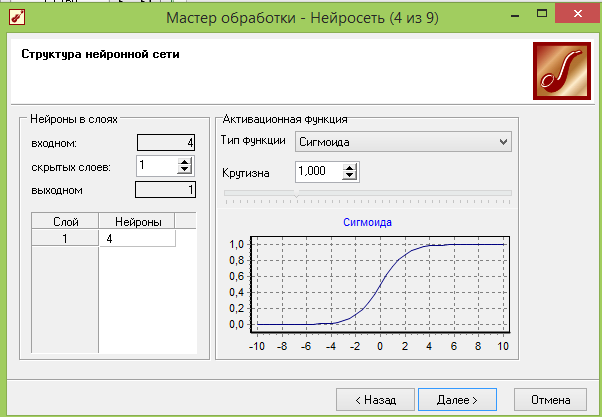


Рис.8 Структура нейронной сети

На 5 –ом шаге выбираем метод обучения НС - Back Prop., на 6-ом - все параметры принимаем по умолчанию, на 7- ом - обучаем НС, нажав «Пуск» (рис.12).

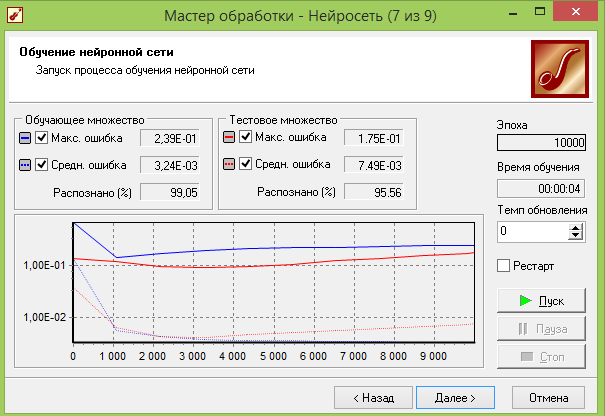


Рис.8 Обучение НС

На 8-ом - указываем те характеристики, которые желательно получить (рис.9).

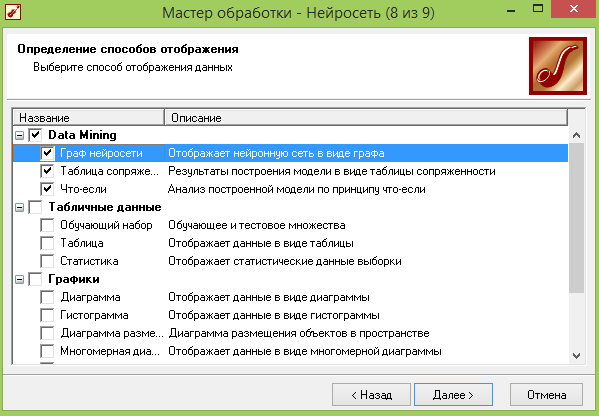


Рис.9 Способы отображения

После этого анализируем итоги (рис.10 -12).

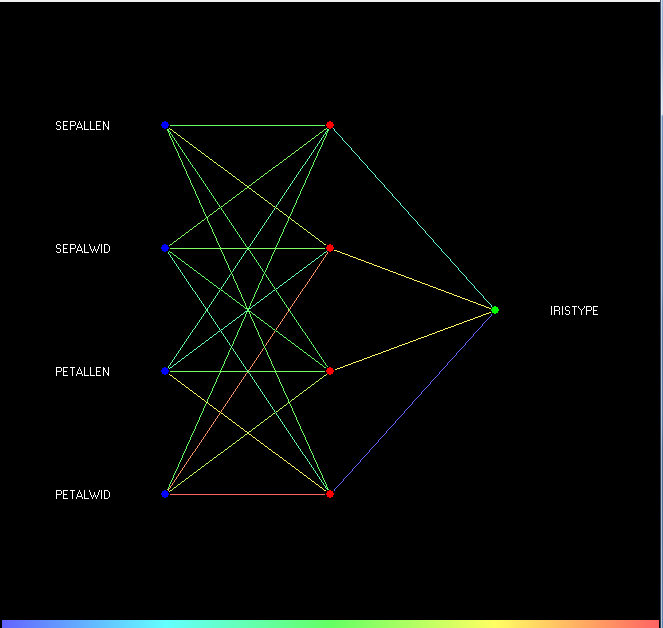


Рис.10 Нейросеть

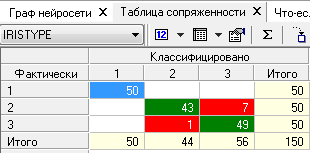


Рис.11 Таблица сопряженности

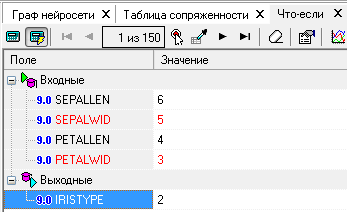


Рис.12 Проверка новых данных

Как видно из рис.10-12, сеть имеет структуру 4-4-1 (как и планировалось); цветы одного типа Setoza - распознаются безошибочно, а другие сорта - с ошибками. С помощью механизма «Что, если» можно определить сорт нового цветка: на рис.12 при вводе новых значений параметров нейросеть определяет тип цветка.

Приложение

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SEPALLEN | SEPALWID | PETALLEN | PETALWID | IRISTYPE |
| 5 | 3,3 | 1,4 | 0,2 | SETOSA |
| 6,4 | 2,8 | 5,6 | 2,2 | VIRGINIC |
| 6,5 | 2,8 | 4,6 | 1,5 | VERSICOL |
| 6,7 | 3,1 | 5,6 | 2,4 | VIRGINIC |
| 6,3 | 2,8 | 5,1 | 1,5 | VIRGINIC |
| 4,6 | 3,4 | 1,4 | 0,3 | SETOSA |
| 6,9 | 3,1 | 5,1 | 2,3 | VIRGINIC |
| 6,2 | 2,2 | 4,5 | 1,5 | VERSICOL |
| 5,9 | 3,2 | 4,8 | 1,8 | VERSICOL |
| 4,6 | 3,6 | 1 | 0,2 | SETOSA |
| 6,1 | 3 | 4,6 | 1,4 | VERSICOL |
| 6 | 2,7 | 5,1 | 1,6 | VERSICOL |
| 6,5 | 3 | 5,2 | 2 | VIRGINIC |
| 5,6 | 2,5 | 3,9 | 1,1 | VERSICOL |
| 6,5 | 3 | 5,5 | 1,8 | VIRGINIC |
| 5,8 | 2,7 | 5,1 | 1,9 | VIRGINIC |
| 6,8 | 3,2 | 5,9 | 2,3 | VIRGINIC |
| 5,1 | 3,3 | 1,7 | 0,5 | SETOSA |
| 5,7 | 2,8 | 4,5 | 1,3 | VERSICOL |
| 6,2 | 3,4 | 5,4 | 2,3 | VIRGINIC |
| 7,7 | 3,8 | 6,7 | 2,2 | VIRGINIC |
| 6,3 | 3,3 | 4,7 | 1,6 | VERSICOL |
| 6,7 | 3,3 | 5,7 | 2,5 | VIRGINIC |
| 7,6 | 3 | 6,6 | 2,1 | VIRGINIC |
| 4,9 | 2,5 | 4,5 | 1,7 | VIRGINIC |
| 5,5 | 3,5 | 1,3 | 0,2 | SETOSA |
| 6,7 | 3 | 5,2 | 2,3 | VIRGINIC |
| 7 | 3,2 | 4,7 | 1,4 | VERSICOL |
| 6,4 | 3,2 | 4,5 | 1,5 | VERSICOL |
| 6,1 | 2,8 | 4 | 1,3 | VERSICOL |
| 4,8 | 3,1 | 1,6 | 0,2 | SETOSA |
| 5,9 | 3 | 5,1 | 1,8 | VIRGINIC |
| 5,5 | 2,4 | 3,8 | 1,1 | VERSICOL |
| 6,3 | 2,5 | 5 | 1,9 | VIRGINIC |
| 6,4 | 3,2 | 5,3 | 2,3 | VIRGINIC |
| 5,2 | 3,4 | 1,4 | 0,2 | SETOSA |
| 4,9 | 3,6 | 1,4 | 0,1 | SETOSA |
| 5,4 | 3 | 4,5 | 1,5 | VERSICOL |
| 7,9 | 3,8 | 6,4 | 2 | VIRGINIC |
| 4,4 | 3,2 | 1,3 | 0,2 | SETOSA |
| 6,7 | 3,3 | 5,7 | 2,1 | VIRGINIC |
| 5 | 3,5 | 1,6 | 0,6 | SETOSA |
| 5,8 | 2,6 | 4 | 1,2 | VERSICOL |
| 4,4 | 3 | 1,3 | 0,2 | SETOSA |
| 7,7 | 2,8 | 6,7 | 2 | VIRGINIC |
| 6,3 | 2,7 | 4,9 | 1,8 | VIRGINIC |
| 4,7 | 3,2 | 1,6 | 0,2 | SETOSA |
| 5,5 | 2,6 | 4,4 | 1,2 | VERSICOL |
| 7,2 | 3,2 | 6 | 1,8 | VIRGINIC |
| 4,8 | 3 | 1,4 | 0,3 | SETOSA |
| 5,1 | 3,8 | 1,6 | 0,2 | SETOSA |
| 6,1 | 3 | 4,9 | 1,8 | VIRGINIC |
| 4,8 | 3,4 | 1,9 | 0,2 | SETOSA |
| 5 | 3 | 1,6 | 0,2 | SETOSA |
| 5 | 3,2 | 1,2 | 0,2 | SETOSA |
| 6,1 | 2,6 | 5,6 | 1,4 | VIRGINIC |
| 6,4 | 2,8 | 5,6 | 2,1 | VIRGINIC |
| 4,3 | 3 | 1,1 | 0,1 | SETOSA |
| 5,8 | 4 | 1,2 | 0,2 | SETOSA |
| 5,1 | 3,8 | 1,9 | 0,4 | SETOSA |
| 6,7 | 3,1 | 4,4 | 1,4 | VERSICOL |
| 6,2 | 2,8 | 4,8 | 1,8 | VIRGINIC |
| 4,9 | 3 | 1,4 | 0,2 | SETOSA |
| 5,1 | 3,5 | 1,4 | 0,2 | SETOSA |
| 5,6 | 3 | 4,5 | 1,5 | VERSICOL |
| 5,8 | 2,7 | 4,1 | 1 | VERSICOL |
| 5 | 3,4 | 1,6 | 0,4 | SETOSA |
| 4,6 | 3,2 | 1,4 | 0,2 | SETOSA |
| 6 | 2,9 | 4,5 | 1,5 | VERSICOL |
| 5,7 | 2,6 | 3,5 | 1 | VERSICOL |
| 5,7 | 4,4 | 1,5 | 0,4 | SETOSA |
| 5 | 3,6 | 1,4 | 0,2 | SETOSA |
| 7,7 | 3 | 6,1 | 2,3 | VIRGINIC |
| 6,3 | 3,4 | 5,6 | 2,4 | VIRGINIC |
| 5,8 | 2,7 | 5,1 | 1,9 | VIRGINIC |
| 5,7 | 2,9 | 4,2 | 1,3 | VERSICOL |
| 7,2 | 3 | 5,8 | 1,6 | VIRGINIC |
| 5,4 | 3,4 | 1,5 | 0,4 | SETOSA |
| 5,2 | 4,1 | 1,5 | 0,1 | SETOSA |
| 7,1 | 3 | 5,9 | 2,1 | VIRGINIC |
| 6,4 | 3,1 | 5,5 | 1,8 | VIRGINIC |
| 6 | 3 | 4,8 | 1,8 | VIRGINIC |
| 6,3 | 2,9 | 5,6 | 1,8 | VIRGINIC |
| 4,9 | 2,4 | 3,3 | 1 | VERSICOL |
| 5,6 | 2,7 | 4,2 | 1,3 | VERSICOL |
| 5,7 | 3 | 4,2 | 1,2 | VERSICOL |
| 5,5 | 4,2 | 1,4 | 0,2 | SETOSA |
| 4,9 | 3,1 | 1,5 | 0,2 | SETOSA |
| 7,7 | 2,6 | 6,9 | 2,3 | VIRGINIC |
| 6 | 2,2 | 5 | 1,5 | VIRGINIC |
| 5,4 | 3,9 | 1,7 | 0,4 | SETOSA |
| 6,6 | 2,9 | 4,6 | 1,3 | VERSICOL |
| 5,2 | 2,7 | 3,9 | 1,4 | VERSICOL |
| 6 | 3,4 | 4,5 | 1,6 | VERSICOL |
| 5 | 3,4 | 1,5 | 0,2 | SETOSA |
| 4,4 | 2,9 | 1,4 | 0,2 | SETOSA |
| 5 | 2 | 3,5 | 1 | VERSICOL |
| 5,5 | 2,4 | 3,7 | 1 | VERSICOL |
| 5,8 | 2,7 | 3,9 | 1,2 | VERSICOL |
| 4,7 | 3,2 | 1,3 | 0,2 | SETOSA |
| 4,6 | 3,1 | 1,5 | 0,2 | SETOSA |
| 6,9 | 3,2 | 5,7 | 2,3 | VIRGINIC |
| 6,2 | 2,9 | 4,3 | 1,3 | VERSICOL |
| 7,4 | 2,8 | 6,1 | 1,9 | VIRGINIC |
| 5,9 | 3 | 4,2 | 1,5 | VERSICOL |
| 5,1 | 3,4 | 1,5 | 0,2 | SETOSA |
| 5 | 3,5 | 1,3 | 0,3 | SETOSA |
| 5,6 | 2,8 | 4,9 | 2 | VIRGINIC |
| 6 | 2,2 | 4 | 1 | VERSICOL |
| 7,3 | 2,9 | 6,3 | 1,8 | VIRGINIC |
| 6,7 | 2,5 | 5,8 | 1,8 | VIRGINIC |
| 4,9 | 3,1 | 1,5 | 0,1 | SETOSA |
| 6,7 | 3,1 | 4,7 | 1,5 | VERSICOL |
| 6,3 | 2,3 | 4,4 | 1,3 | VERSICOL |
| 5,4 | 3,7 | 1,5 | 0,2 | SETOSA |
| 5,6 | 3 | 4,1 | 1,3 | VERSICOL |
| 6,3 | 2,5 | 4,9 | 1,5 | VERSICOL |
| 6,1 | 2,8 | 4,7 | 1,2 | VERSICOL |
| 6,4 | 2,9 | 4,3 | 1,3 | VERSICOL |
| 5,1 | 2,5 | 3 | 1,1 | VERSICOL |
| 5,7 | 2,8 | 4,1 | 1,3 | VERSICOL |
| 6,5 | 3 | 5,8 | 2,2 | VIRGINIC |
| 6,9 | 3,1 | 5,4 | 2,1 | VIRGINIC |
| 5,4 | 3,9 | 1,3 | 0,4 | SETOSA |
| 5,1 | 3,5 | 1,4 | 0,3 | SETOSA |
| 7,2 | 3,6 | 6,1 | 2,5 | VIRGINIC |
| 6,5 | 3,2 | 5,1 | 2 | VIRGINIC |
| 6,1 | 2,9 | 4,7 | 1,4 | VERSICOL |
| 5,6 | 2,9 | 3,6 | 1,3 | VERSICOL |
| 6,9 | 3,1 | 4,9 | 1,5 | VERSICOL |
| 6,4 | 2,7 | 5,3 | 1,9 | VIRGINIC |
| 6,8 | 3 | 5,5 | 2,1 | VIRGINIC |
| 5,5 | 2,5 | 4 | 1,3 | VERSICOL |
| 4,8 | 3,4 | 1,6 | 0,2 | SETOSA |
| 4,8 | 3 | 1,4 | 0,1 | SETOSA |
| 4,5 | 2,3 | 1,3 | 0,3 | SETOSA |
| 5,7 | 2,5 | 5 | 2 | VIRGINIC |
| 5,7 | 3,8 | 1,7 | 0,3 | SETOSA |
| 5,1 | 3,8 | 1,5 | 0,3 | SETOSA |
| 5,5 | 2,3 | 4 | 1,3 | VERSICOL |
| 6,6 | 3 | 4,4 | 1,4 | VERSICOL |
| 6,8 | 2,8 | 4,8 | 1,4 | VERSICOL |
| 5,4 | 3,4 | 1,7 | 0,2 | SETOSA |
| 5,1 | 3,7 | 1,5 | 0,4 | SETOSA |
| 5,2 | 3,5 | 1,5 | 0,2 | SETOSA |
| 5,8 | 2,8 | 5,1 | 2,4 | VIRGINIC |
| 6,7 | 3 | 5 | 1,7 | VERSICOL |
| 6,3 | 3,3 | 6 | 2,5 | VIRGINIC |
| 5,3 | 3,7 | 1,5 | 0,2 | SETOSA |
| 5 | 2,3 | 3,3 | 1 | VERSICOL |

**Задание**

**1. Создать базу примеров**

Все признаки во всех вариантах разыгрываются, исходя из равномерного закона распределения.

*Вариант 1 Подбор персонала*

Примем, что фирма занимается, в основном, разработкой программного обеспечения. Тогда в качестве характеристик, которыми должен обладать кандидат на вакантное место в этой фирме, выберем следующие: *Х*1 - возраст (*Age*); *Х*2 - образование (*Edu*); *Х*3 - опыт (*Exp*); *Х*4 - знание иностранного языка (*Eng*); *Х*5 - обладание специальными навыками в программировании (*Spe*); *Х*6 - умение работать в команде (*Com*).

Разыгрывается 2 класса: годен – не годен; разыграть по 20 субъектов каждого класса;

Х1 - возраст (годы): класс 1 – от 22 до 35 лет, класс 2 – от 35 до 45 лет;

Х2 - образование (балльная оценка): кл.1- от 6 до 10; кл.2 - (от 1 до 6);

Х3 -опыт (годы): кл.1 - от 3 до 6; кл2 - от 1 до 3;

Х4 - ин.язык (баллы): кл.1- от 6 до 10; кл.2 от 2 до 7;

Х5 - навык програм. (баллы): кл.1- (6-10); кл.2 - (1-6);

Х6 - работа в ком. (баллы): кл.1- (7-10); кл.2 - (1-6);

Вся таблица: 40 строк, 7 столбцов (6 признаков + название класса).

Пример фрагмента таблицы

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Х1 | Х2 | Х3 | Х4 | Х5 | 1Х6 | Класс |
| 23 | 7 | 4 | 9 | 8 | 9 | не годен |
| 30 | 6 | 7 | 7 | 6 | 8 | годен |
| 20 | 1 | 1 | 2 | 2 | 1 | не годен |
| 27 | 6 | 7 | 9 | 9 | 9 | годен |
| 33 | 10 | 7 | 9 | 6 | 8 | годен |

*Вариант 2 Компетенции персонала*

В качестве входных параметров были взяты оценки компетенции, выставленные руководителем, подчиненными, коллегами и, кроме того, была включена самооценка. Таким образом, число входов равно четырем. Диапазон шкалы оценок: от 1 до 10 баллов.

Выходной параметр - класс сотрудников. Установим для них два значения: с высокими (класс 1) и низкими (класс 2) оценками и смоделируем базу примеров, используя метод Монте-Карло. Объем моделирования определим в 40 сотрудников, по 20 человек для каждого класса. В итоге, база примеров состоит из 40 строк и 5 столбцов (4 столбца - входные признаки, пятый - номер класса), т.е. каждый сотрудник описывается вектором-строкой из четырех компонентов.

Пример фрагмента таблицы

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X1 | X2 | X3 | X4 | Class |
| 3 | 4 | 2 | 5 | 0 |
| 7 | 6 | 8 | 9 | 1 |
| 3 | 3 | 5 | 2 | 0 |
| 5 | 6 | 6 | 4 | 0 |
| 10 | 6 | 10 | 5 | 1 |
| 10 | 8 | 8 | 9 | 1 |

*Вариант 3 Выбор стратегии*

Для конкретности решения допустим, что предприятие выпускает наукоемкую продукцию, в частности, комплекс средств контроля работы технического оборудования. Основываясь на опыте решения подобных задач в области стратегического менеджмента, в качестве показателей, от которых зависит стратегия предприятия, выберем следующие: *Х*1 - частота инноваций продукции; *Х*2 - уровень издержек при обновлении продукции;  *Х*3 - уровень качества продукции; *Х*4 - контроль над каналами распределения продукции; *Х5* - уровень технической поддержки продукции.

Каждый признак имеет балльную шкалу: от 0 до 10. Всего 3 вида стратегии: первая, вторая и третья; 10 вариантов розыгрыша в каждом классе.

Принять: первый вид стратегии - самый лучший, второй - средний и третий - хуже среднего.

Например, признак Х1 имеет такие границы:

1 кл - (7-10); 2 кл. - (4-7); 2 кл. - (1-4).

Аналогично признаки Х3 и Х4.

Признаки Х2 и Х5 имеют такие границы:

1 кл - (1-5); 2 кл - (4-7); 3 кл. - (6-10)

В этом варианте обучающая выборка состоит из первоначальных 30 строк и 6 столбцов: пять столбцов определяют отобранные показатели; шестой - класс стратегии (*first, second, third*).

Пример фрагмента таблицы

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X1 | X2 | X3 | X4 | X5 | Class |
| 7,2 | 7,38 | 3,55 | 4,94 | 3,07 | 1,00 |
| 17,7 | 7,10 | 3,36 | 5,41 | 3,13 | 1,00 |
| 10,5 | 7,60 | 3,99 | 3,40 | 3,93 | 1,00 |
| 4,3 | 7,90 | 3,21 | 5,92 | 3,36 | 1,00 |

*Вариант 4 Конкурентоспособность предприятия*

В качестве факторов, влияющих на КСП предприятия, примем следующие:

1. качество выпускаемой продукции;
2. товарно-сбытовая деятельность предприятия;
3. организационно-техническое совершенство производства;
4. финансовое состояние;
5. имидж предприятия.

Такой набор факторов является достаточным для целей исследования, так как он охватывает основные аспекты КСП и не противоречит общепринятым тенденциям.

Классы КСП организации установим такими:

1. нормальная (*normal*) – результаты конкретных взаимодействий совпадают с ожиданиями участвующих заинтересованных сторон;
2. ниже нормальной (*less*)  – фактические результаты не соответствуют ожиданиям. Затем заинтересованные стороны, вовлеченные в эту ситуацию, предпримут действия, чтобы отказаться от взаимодействия с компанией и перейти к другой, более привлекательной.
3. выше нормальной (*more*) – фактические результаты выше, чем ожидалось. Заинтересованные стороны, у которых есть основания для таких оценок, стремятся укрепить свои отношения с компанией.

Всего 3 класса, по 10 вариантов в каждом классе. Оценка всех признаков - балльная: от 0 до 10 баллов.

Пример фрагмента таблицы

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Х1 | Х2 | Х3 | Х4 | Х5 | Класс |
| 1,1 | 2,2 | 2,3 | 0,0 | 1,4 | Less |
| 0,3 | 0,2 | 0,7 | 1,8 | 1,9 | Less |
| 1,8 | 2,3 | 1,3 | 1,8 | 0,1 | Less |
| 2,7 | 2,4 | 0,7 | 1,3 | 2,3 | Less |
| 2,7 | 2,6 | 2,6 | 1,3 | 1,6 | Less |
| 2,9 | 2,9 | 1,5 | 1,8 | 1,3 | Less |
| 0,0 | 0,4 | 3,0 | 1,9 | 2,5 | Less |
| 1,2 | 1,8 | 1,0 | 3,0 | 1,0 | Less |
| 2,6 | 2,3 | 0,3 | 2,5 | 0,3 | Less |
| 0,4 | 1,4 | 0,2 | 2,6 | 2,5 | Less |
| 6,2 | 3,6 | 3,4 | 5,6 | 3,5 | Normal |
| 4,1 | 6,5 | 3,7 | 3,3 | 6,1 | Normal |

**2. Создание нейронной сети**

Используя разыгранные таблицы, создать текстовый файл, сохранить его, сконструировать трехслойный персептрон в системе *Deductor*, обучить сеть, проверить ее работоспособность с помощью процедуры "Что, если".

**Вопросы**

1. Что имитируют нейронные сети (НС)?
2. Для решения каких задач применяют НС?
3. Как определяется число слоев и нейронов в них при решении задачи классификации?
4. Какие методы обучения НС существуют?
5. В чем сущность метода обратного распространения ошибки?
6. Нарисовать схему НС для классификации трех видов объектов (стол, стул, шкаф), каждый из которых характеризуется длиной, шириной, высотой, массой?