**Лабораторная работа №3**

**"Нейронные сети. Метод backpropagation"**

1. Решить задачу приближения простого набора данных. Значения взять

из папки simplefit\_dataset.

2. Задача - по выбору. В задаче необходимо разделить исходные данные на

контрольную и обучающую выборки (примерно 80-20%). Оценивать свой

результат по работе сети на контрольной выборке. Данные в основном взяты

с сайта UCI Machine Learning Repository. Можете выбрать себе задачу прямо

оттуда.

Варианты задачи 2:

1. Решить задачу оценки средней цены дома.

https://archive.ics.uci.edu/ml/datasets/Housing

Данные в папке house\_dataset. Имеем 506 домов с известной ценой и 13

входными признаками. Количество признаков при желании можно немного

уменьшить.

1. Per capita crime rate per town (уровень престуности на душу населения)

2. Proportion of residential land zoned for lots over 25,000 sq. ft.

3. proportion of non-retail business acres per town

4. 1 if tract bounds Charles river, 0 otherwise

5. Nitric oxides concentration (parts per 10 million)

6. Average number of rooms per dwelling

7. Proportion of owner-occupied units built prior to 1940

8. Weighted distances to five Boston employment centres

9. Index of accessibility to radial highways

10. Full-value property-tax rate per $10,000

11. Pupil-teacher ratio by town

12. 1000(Bk - 0.63)^2б where Bk is the proportion of blacks by town

13. Percent lower status of the population

2. Решить задачу подсчета колец раковины моллюска морское ушко.

https://archive.ics.uci.edu/ml/datasets/Abalone

Возраст моллюска определяется путем разрезания его раковины,

окрашивания и подсчета количества колец с помощью микроскопа -- скучно

и трудоемко. В связи с этим актуальна задача предсказания возраста

моллюска по другим измерениям.

Данные 4177 моллюсков в папке abalone\_dataset. Оцениваются 8 показателей.

Выходной вектор -- количество колец.

1. Sex: M, F, and I (infant)

2. Length

3. Diameter

4. Height

5. Whole weight

6. Shucked weight

7. Viscera weight

8. Shell weight

3. Решить задачу оценки процента жира в организме.

http://lib.stat.cmu.edu/datasets/bodyfat

Имеем 252 человека с 13 измеренными показателями:

1. Age (years)

2. Weight (lbs)

3. Height (inches)

4. Neck circumference (cm)

5. Chest circumference (cm)

6. Abdomen 2 circumference (cm)

7. Hip circumference (cm)

8. Thigh circumference (cm)

9. Knee circumference (cm)

10. Ankle circumference (cm)

11. Biceps (extended) circumference (cm)

12. Forearm circumference (cm)

13. Wrist circumference (cm)

Выходной сигнал - процент жира.

4. Chemical Dataset– Оценка сигнала датчика по сигналам восьми других

датчиков;

Имеем 498 измерений с 8 датчиков во время химического процесса.

Возможность оценить по ним показания девятого датчика позволит удалить

его. Поэтому задача оценки его параметров по данным 8 других датчиков

актуальна.

5. Engine Dataset– Оценка крутящего момента и выбросов двигатели от

использованного топлива и скорости

2199 двигателей, 2 входа (топливо, скорость), 2 выхода (крутящий момент,

выбросы оксида азота)

6. Ирисы Фишера (iris\_dataset) . Для каждого экземпляра ириса известны 4

величины: длина чашелистика, ширина чашелистика, длина лепестка,

ширина лепестка. Задача состоит в классификации данных (три группы).

7. Breast Cancer – классификация опухоли как доброкачественной либо

злокачественной в зависимости от характеристик образцов биопсии.

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original)

Данные в cancer\_dataset. 9 входных признаков:

1. Clump thickness

2. Uniformity of cell size

3. Uniformity of cell shape

4. Marginal Adhesion

5. Single epithelial cell size

6. Bare nuclei

7. Bland chomatin

8. Normal nucleoli

9. Mitoses

и 2 выхода (доброкачественная/злокачественная)

8. Классификация стекол - оконное/неоконное в зависимости от химического

состава стекла. Имеется 214 образцов стекла, 9 входных признаков:

1. Показатель преломления

2. Натрий (единица измерения: весовой процент в соответствующий оксид)

3. Магний

4. Алюминий

5. Кремний

6. Калий

7. Кальций

8. бария

9. Железо

и матрица 2х214, указывающая правильную категорию стекла. Данные в

glass\_dataset.

9. Классификация раков по полу М/Ж. Имеется 200 крабов, 6 входных

признаков

1. Species

2. Frontal lip

3. Rearwidth

4. Length

5. Width

6. Depth.

и выходная матрица 2х200 с правильным ответом. Данные в crab\_dataset.

10. Распознавание рукописных цифр. Смотрите каталог MNIST. В базе

содержатся 5000 изображений рукописных цифр в градациях серого,

размером 20х20 пикселей, "развернутых" в вектор из 400 элементов. Также в

базе находится вектор из 5000 элементов, содержащий правильный выход

(цифру). Для совместимости с Матлабом, в котором нет нулевого индекса, 0

всюду заменен цифрой 10.

Итак, у нас 400 входов и 10 выходов. Количество нейронов скрытого слоя

выбираем сами.