

1 Лабораторная работа №1 – Классификация рукописных цифр с использованием метода ближайшего соседа

Цель: Реализовать в системе Matlab классификатор рукописных цифр на основании метода ближайшего соседа.

1.1 Теоретические сведения

1.1.1 База изображений

В лабораторной работе необходимо построить в системе Matlab классификатор рукописных цифр. Для обучения классификатор имеется подготовленная база изображений хранящаяся в файле "data.mat". Для загрузки базы в систему Matlab необходимо использовать команду

```
load('data');
```

Данная база состоит из 3 структур: training, validation, test. Каждая из структур включает в себя массивы inputs (исходные изображения) и targets (где храниться метка класса, к которому относится соответствующее изображение). Для обучения классификатора необходимо использовать обучающую базу training, а для проверки и оценки производительности базу test. Каждая цифра хранится в базе в виде черно-белого изображения размеров 16x16 пикселей, но которое для удобства хранения представлено в виде вектора 1x256. Для того, чтобы увидеть информацию по конкретному изображению кон цифры из базы нужно выполнить следующий код (вывод на экран 4 элемента из базы):

```
tr_base=data.training.inputs;  
tr_trg =data.training.targets;  
figure;  
subplot(1,5,[1 2])  
plot(data.training.inputs(:,4),'-o'); xlim([1 256]);  
title('Входной вектор (Inputs)');  
subplot(1,5,3);  
pcolor(rot90(reshape(tr_base(:,4),16,16))); shading flat;  
colormap gray;  
title('Изображение');  
subplot(1,5,[4 5]);  
stem(data.training.targets(:,4),'r-o'); xlim([1 10]);  
title('Целевые значения (targets)');
```

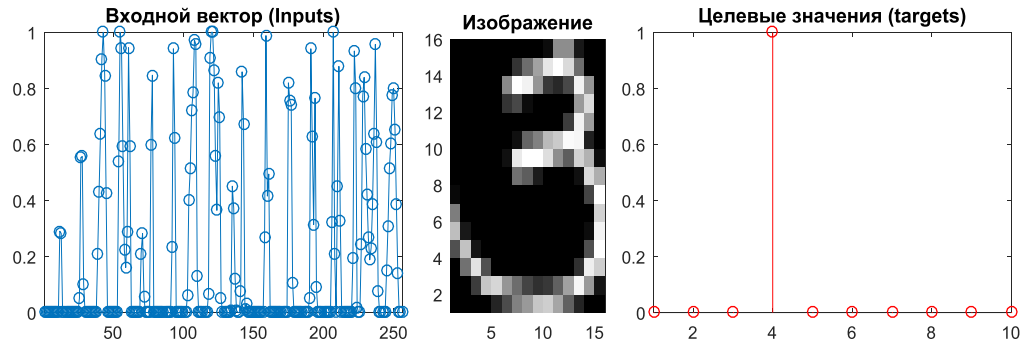


Рисунок 1 – Исходные данные

1.1.2 Метод ближайшего соседа

Метод ближайшего соседа один из самых простых и эффективных методов классификации. Он требует задания на обучающем множестве элементов функции расстояния $\phi(\mathbf{x}_i, \mathbf{x}_j)$, которая определяет сходство между элементами (векторами) \mathbf{x}_i и \mathbf{x}_j . Классифицируемый элемент относится к тому классу, которому принадлежат ближайшие к нему элементы обучающей выборки.

1.1.3 Мера близости

Между двумя векторами можно предложить различные функции расстояния. Ниже перечисленные наиболее распространенные из них:

Евклидово расстояние:

$$D_E(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x}_1 - \mathbf{y}_1)^2 + \dots + (\mathbf{x}_r - \mathbf{y}_r)^2}. \quad (1.1)$$

Модульное расстояние:

$$D_M(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^r |\mathbf{x}_i - \mathbf{y}_i|. \quad (1.2)$$

l_p -норма:

$$D_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^r |\mathbf{x}_i - \mathbf{y}_i|^p \right)^{1/p} \quad (1.3)$$

Мера Джеффриса-Матуситы:

$$D_{JM}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^r (\sqrt{\mathbf{x}_i} - \sqrt{\mathbf{y}_i})^2 \quad (1.4)$$

"Коэффициент дивергенции":

$$D_{CD}(\mathbf{x}, \mathbf{y}) = \left(\frac{1}{r} \sum_{i=1}^r \left(\frac{\mathbf{x}_i - \mathbf{y}_i}{\mathbf{x}_i + \mathbf{y}_i} \right)^2 \right)^{1/2} \quad (1.5)$$

Расстояние на основе коэффициента корреляции:

$$D_{Corr}(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_{i=1}^r \mathbf{x}_i \mathbf{y}_i}{(\sum_{i=1}^r \mathbf{x}_i^2 \sum_{i=1}^r \mathbf{y}_i^2)^{1/2}} \quad (1.6)$$

1.1.4 Матрица спутывания

Матрица спутывания позволяет оценить производительность работы классификатора. Для классификатора на M классов матрица спутывания будет иметь размер $M \times M$. В ячейке с индексом (i, j) хранится значение вероятности отнесения классификатором объекта i -го класса к классу j .

1.2 Задание

1. Постройте классификатор рукописных цифр на основе метода ближайшего соседа. Функция расстояния выбирается в соответствии таблицей

Таблица 1 – Варианты заданий

Номер варианта	Функция расстояния
1	Используйте выражение (1.1).
2	Используйте выражение (1.2).
3	Используйте выражение (1.3) с параметром $p = 3$.
4	Используйте выражение (1.4).
5	Используйте выражение (1.5).
6	Используйте выражение (1.3) с параметром $p = 5$.
7	Используйте выражение (1.6).
8	Используйте выражение (1.3) с параметром $p = 4$.

2. Для полученного классификатора постройте матрицу спутывания.

1.3 Порядок выполнения работы

1. Выполнить практическое задание.
2. Продемонстрировать результат работы преподавателю.
3. Оформить отчет. В отчете привести исходных код и кратко изложить решение.
4. Защитить работу