МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования   
**«Национальный исследовательский   
Нижегородский государственный университет им. Н.И. Лобачевского»**

**(ННГУ)**

**Институт информационных технологий, математики и механики**

Направление подготовки: «Прикладная математика и информатика»

Магистерская программа: «Вычислительные методы и суперкомпьютерные технологии»

Образовательный курс «Методы глубокого обучения для решения задач компьютерного зрения»

**ОТЧЕТ**

по лабораторной работе №1

**«Реализация метода обратного распространения ошибки для двуслойной полностью связанной нейронной сети»**

**Выполнил:**

студент группы 381603м4

Уваров Денис Вадимович

Нижний Новгород

2017

**Оглавление**

[Постановка задачи 3](#_Toc501712192)

[Метод обратного распространения ошибки 4](#_Toc501712193)

[Общая схема метода обратного распространения ошибки 5](#_Toc501712194)

[Вывод математических формул 6](#_Toc501712195)

[Программная реализация 8](#_Toc501712196)

[Результаты экспериментов 9](#_Toc501712197)

# Постановка задачи

Целью данной лабораторной работы является изучение метода обратного распространения ошибки для обучения многослойных искусственных нейронных сетей и его реализация на примере полностью связанной нейронной сети.

В качестве демонстрации работоспособности реализованного метода будет решена задача классификации рукописных цифр из набора данных MNIST с помощью полностью связанной нейронной сети с одним скрытым слоем.

В процессе выполнения лабораторной работы будут решены следующие задачи:

1. Изучение общей схемы метода обратного распространения ошибки;
2. Вывод математических формул для всех этапов метода;
3. Разработка программной реализации метода обратного распространения ошибки;
4. Реализация демонстрационного приложения для классификации рукописных цифр с помощью разработанной программной реализации.

# Метод обратного распространения ошибки

Пусть у искусственной нейронной сети есть множество входов , множество выходов и множество внутренних узлов, расположенных на скрытых слоях. Вес связи соединяющей 𝑖-ый и 𝑗-ый узлы обозначим через , а через – выход 𝑖-го узла.

В качестве функции ошибки нейронной сети для задачи классификации рассматривается кросс-энтропия:

, (1)

где:

* 𝐿 – количество примеров в обучающей выборке,
* 𝑀 – количество классов и нейронов выходном слое,
* – метка входного примера из обучающей выборки,
* – выход нейронной сети для того же примера.

Задача обучения нейронной сети может быть сформулирована следующим образом:

(2)

Задача (2) может быть решена численными методами, например, методом градиентного спуска или квазиньютоновскими методами.

В данной лабораторной работе рассматривается последовательный (стохастический) режим обучения нейронной сети. В таком режиме корректировка весов выполняется после каждого предъявления примера из обучающей выборки.

В качестве критерия остановки можно использовать фиксированное число предъявлений всей обучаемой выборки (максимально допустимое количество эпох), достигнутая точность поиска минимума функции потерь, либо слабое изменение весов.

# Общая схема метода обратного распространения ошибки

Общая схема метода обратного распространения ошибки может быть представлена следующим образом:

1. Инициализация весов и установка параметров нейронной сети.
2. Пока не выполнится один из критериев остановки ̅̅̅̅ (достигнуто максимальное число эпох или достигнутая точность обучения):
   1. Прямой проход нейронной сети
      1. Подаем входные параметры системы и вычисляем значения выходных сигналов нейронов поочередно на каждом из слоев.
   2. Обратный проход нейронной сети
      1. Вычисляем ошибку на каждом из слоев
      2. Делаем корректировку весов
3. Полученные значения весов 𝜔 использовать для дальнейшей классификации.

# Вывод математических формул

Как видно из схемы метода обратного распространения ошибки, описанном в предыдущей секции, необходимым условием реализации и применения данного метода является знание частных производных и градиентов функции потерь на всех слоях нейронной сети.

Выход , -го нейрона входного слоя:

(3)

где – значение 𝑗-ой компоненты входного вектора, что в данной лабораторной работе является интенсивностью 𝑗−го пикселя.

Выход  𝑗-го нейрона скрытого слоя вычисляется следующим образом:

, (4)

где − функция активации слоя.

В данной лабораторной работе в качестве функции активации скрытых слоёв рассматривается сигмоидальная функция

, (5)

где - параметр функции активации.

Выход , -го нейрона выходного слоя вычисляется аналогично (4), только в качестве функции активации рассматривается функция Softmax:

. (6)

Формулы (4) – (6) позволяют вычислить выход нейронной сети, а именно выполнить прямой ход алгоритма обратного распространения ошибки

Перед началом обратного хода необходимо вычислить градиент функции ошибки (1) по сумматору нейронов выходного слоя с функцией активации (6) для одного предъявленного примера из обучающей выборки при условии .

(7)

где .

Ошибка в узлах скрытого слоя является произведением взвешенной ошибки на последующем слое на производную функции активации этого слоя. Производные выбранных функций активации (5) – (6) могут быть вычислены с использованием значений самой функции. Таким образом в случае сигмоидальной функции активации (5) получаем:

(8)

Подстановка полученных формул (7) и (8) в общую схему метода обратного распространения ошибки позволяет вычислить ошибки на всех слоях нейронной сети и произвести коррекцию весом для минимизации функции ошибки.

Коррекция весов производится по следующей формуле:

(9)

где:

* 𝜂 – скорость обучения,
* *r* – порядковый номер эпохи.

# Программная реализация

Программная реализация метода обратного распространения ошибки была разработана на языке в соответствии с приведённым выше алгоритмом и выведенными формулами для коррекции весов.

Вся реализация представлена в одном файле, в котором, для большей наглядности, было произведено разбиение на логические блоки алгоритма.

# Результаты экспериментов

Точность классификации определялась согласно формуле

(10)

При проведении экспериментов в качестве критерия остановки использовался комбинированный подход – максимальное количество эпох (), и порог достигнутой точности на валидационной выборке ().

Начальные веса инициализируются случайными значениями из равномерного распределения с границами 0.003 и 0.007.

**Результаты проведенных экспериментов:**

|  |  |  |  |
| --- | --- | --- | --- |
| Количество нейронов на скрытом слое | Скорость обучения скрытого слоя | Количество эпох | Достигнутая точность на тестовой выборке |
| **100** | **0.05** | **20** | **0.9762** |
| **150** | **0.05** | **20** | **0.9771** |
| **100** | **0.1** | **20** | **0.9721** |
| **50** | **0.05** | **20** | **0.9660** |
| **200** | **0.1** | **20** | **0.9750** |