



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н. Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н. Э. Баумана)

---

ФАКУЛЬТЕТ «Информатика, искусственный интеллект и системы управления»

---

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

---

## ОТЧЕТ

по лабораторной работе № 4  
по курсу «Основы искусственного интеллекта»  
на тему: «Базовая искусственная нейронная сеть»

Студент ИУ7-13М  
(Группа)

\_\_\_\_\_  
(Подпись, дата)

Орду М. А.  
(И. О. Фамилия)

Преподаватель

\_\_\_\_\_  
(Подпись, дата)

Строганов Ю. В.  
(И. О. Фамилия)

2025 г.

# СОДЕРЖАНИЕ

<b>1</b>	<b>Теоретический раздел . . . . .</b>	<b>3</b>
1.1	Цель работы . . . . .	3
1.2	Задачи работы . . . . .	3
1.3	Нейронные сети и функции активации . . . . .	3
1.4	Неравенство Чебышева для оценки минимального размера выборки . . . . .	4
<b>2</b>	<b>Практический раздел . . . . .</b>	<b>6</b>
2.1	Программная реализация . . . . .	6
2.2	Результаты экспериментов . . . . .	6
	<b>ПРИЛОЖЕНИЕ А . . . . .</b>	<b>8</b>

# 1 Теоретический раздел

## 1.1 Цель работы

Целью данной лабораторной работы является изучение методов классификации изображений с помощью нейронных сетей на примере датасета MNIST, а также исследование влияния архитектуры сети и соотношения обучающей и тестовой выборок на качество обучения.

## 1.2 Задачи работы

1. Создать нейронные сети с различным количеством скрытых слоев (0, 1, 5) с использованием активации ReLU и функции потерь KL Divergence.
2. Обучить сети на различных долях обучающей выборки (10%, 20%, ..., 90%) и оценить точность на обучающей и тестовой выборках.
3. Определить состояния переобучения и недообучения.
4. Рассчитать минимальный размер обучающей выборки с помощью неравенства Чебышёва.

## 1.3 Нейронные сети и функции активации

Нейронная сеть представляет собой последовательность слоёв, каждый из которых выполняет линейное преобразование и нелинейную активацию. Для скрытых слоёв использовалась активация ReLU (*англ. Rectified Linear Unit*), график которой представлен на рисунке 1.1:

$$f(x) = \max(0, x) \quad (1.1)$$

На выходном слое применялся LogSoftmax для использования с функцией потерь KL Divergence:

$$\text{LogSoftmax}(z_i) = \log \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (1.2)$$

Функция потерь KL Divergence между предсказанным распределени-

ем  $q$  и истинным распределением  $p$  записывается как:

$$D_{KL}(p \parallel q) = \sum_i p_i \log \frac{p_i}{q_i} \quad (1.3)$$

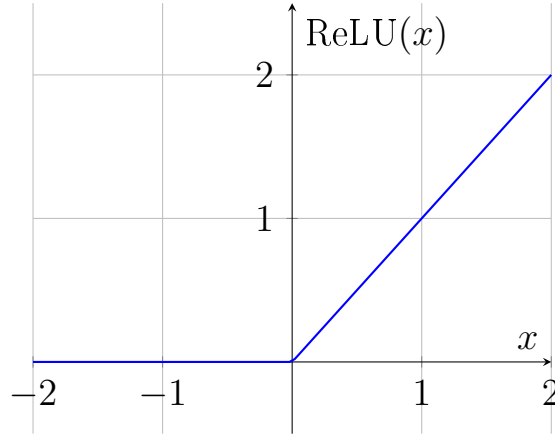


Рисунок 1.1 – График функции ReLU

## 1.4 Неравенство Чебышева для оценки минимального размера выборки

Для случайной величины  $X$  с математическим ожиданием  $\mu$  и дисперсией  $\sigma^2$  справедливо неравенство Чебышева:

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2} \quad (1.4)$$

где  $\bar{X}$  — среднее по  $N$  наблюдениям,  $\epsilon$  — допустимая ошибка.

При оценке минимального размера обучающей выборки для задачи классификации удобно использовать дисперсию Бернулли:

$$\sigma^2 = p(1 - p) \quad (1.5)$$

где  $p$  — ожидаемая точность модели на выборке;  $(1 - p)$  — вероятность ошибки классификации.

Тогда неравенство Чебышева примет вид:

$$P(|\bar{X} - p| \geq \epsilon) \leq \frac{p(1 - p)}{N\epsilon^2} \quad (1.6)$$

Из него выводим минимальный размер выборки:

$$N \geq \frac{p(1-p)}{\delta \epsilon^2} \tag{1.7}$$

где  $\delta = 1 - P$  — вероятность того, что отклонение превысит  $\epsilon$ .

## 2 Практический раздел

### 2.1 Программная реализация

Программная реализация выполнялась на языке Python с использованием библиотеки PyTorch.

### 2.2 Результаты экспериментов

При увеличении доли обучающей выборки точность модели растёт. Сети с большим количеством скрытых слоёв демонстрируют более высокую точность на `train` выборке, что может приводить к переобучению.

Для оценки минимального размера выборки использовалось неравенство Чебышёва, что позволило определить необходимое количество данных для гарантированного достижения заданной точности.

## ВЫВОДЫ

1. Архитектура сети и количество скрытых слоёв напрямую влияют на точность классификации и риск переобучения.
2. Увеличение доли обучающей выборки повышает обобщающую способность модели.
3. Минимальный размер выборки для гарантированного качества можно рассчитать с помощью неравенства Чебышёва.
4. На практике для MNIST сеть с одним или несколькими скрытыми слоями и ReLU активацией достигает высокой точности даже на ограниченной выборке.

## ПРИЛОЖЕНИЕ А