

Исследование процесса распознавания рукописных цифр полносвязной двухслойной нейронной сетью

М. И. Попов, e-mail: mihail_semilov@mail.ru

Воронежский государственный университет

Аннотация. Рассмотрена задача распознавания рукописных цифр с помощью полносвязной двухслойной нейронной сети. В качестве обучающей и тестовой выборки использован набор MNIST. Оценена сложность распознавания каждой цифры и схожесть цифр между собой. Выявлена зависимость между количеством распознаваемых классов изображения, размером скрытого слоя и точностью распознавания.

Ключевые слова: полносвязная нейронная сеть, распознавание рукописных цифр, MNIST

Введение

В современном мире нейронные сети находят все больше областей применения. Классической задачей, решаемой нейронной сетью традиционно считается распознавание изображений. Данная задача относится к задаче классификации. Сюда же можно отнести и распознавание речи, с чем также хорошо справляются нейронные сети. Другим классом задач являются задачи прогнозирования. Например, экономические прогнозы на уровень спроса, объем продаж, рост или падение акций. Сюда же можно отнести и анализ платёжеспособности клиента и риска предоставления ему кредита, и анализ надёжности фирмы и измерение вероятности её банкротства. Без ИНС трудно себе представить и обработку запроса в поисковых системах интернета. Отдельно следует отметить использование ИНС в медицине. Они становятся незаменимыми при постановке диагнозов многих заболеваний. Несмотря на повсеместное внедрение нейронных сетей, выбор архитектуры сети и оптимизация ее гиперпараметров представляют собой довольно трудоемкую задачу, решаемую, как правило, методами перебора [1,2]. В данной работе проведено эмпирическое исследование процесса распознавания рукописных цифр полносвязной двухслойной нейронной сетью.

1. Исследование процесса распознавания рукописных цифр

Рассмотрим полносвязную двухслойную нейронную сеть [3]. Будем обучать ее на тренировочном наборе базы данных MNIST. В качестве функции ошибки выберем среднюю квадратичную ошибку (MSE). Оптимизировать веса будем с помощью стохастического градиентного спуска (SGD)[4]. В качестве функции активации в скрытом слое выберем гиперболический тангенс, а в выходном softmax. Скорость обучения (learning rate) положим равной 0,005 [5]. Проанализируем, как ошибки распознавания распределяются между цифрами. На рис. 1 изображен график процента ошибок при распознавании каждой цифры.

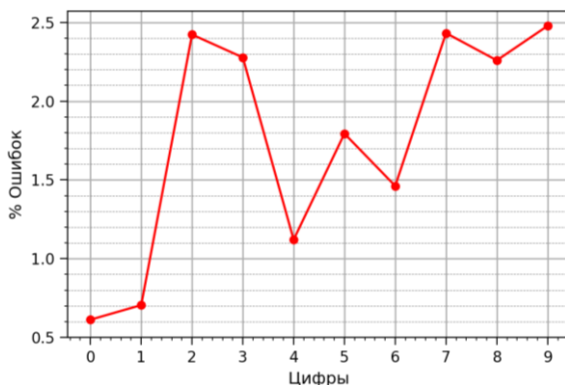


Рис. 1. Процент ошибок при распознавании цифр.

В вычислительном эксперименте размер скрытого слоя определим в 300 нейронов. При данных параметрах длительность обучения до наивысшей точности составит 95 эпох. В базе данных MNIST количество различных цифр не одинаково, поэтому количество ошибок берется по отношению к количеству соответствующих цифр в наборе данных. По графику видно, наибольшее число ошибок достигается при распознавании цифры 9, а наименьшее при определении 0. Расположим цифры в порядке возрастания сложности: 0, 1, 4, 6, 5, 8, 3, 2, 7, 9.

Далее проанализируем, какие цифры чаще всего предлагает сеть в случае ошибки. На рис. 2 представлен график частоты цифр в нераспознанных примерах. По-прежнему лидирует цифра 9, ее не только тяжело распознать, она также чаще всего предлагается сетью, в качестве ошибочного прогноза. Неожиданно 2 место заняла 4, не смотря на то, что она достаточно легко распознается. С наименьшей ошибкой прогнозируется цифра 1. Заметим, что на тренировочном наборе, сеть не ошибается ни разу.

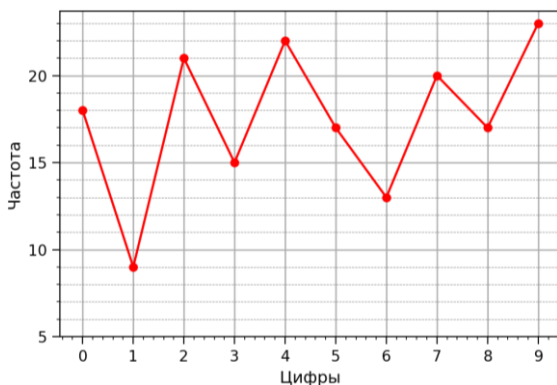


Рис. 2. Частота цифр в нераспознанных изображениях.

Анализ графиков 1-2 позволяет оценить сложность распознавания цифр. Расположим цифры в порядке возрастания сложности: 1, 0, 6, 5, 4, 2, 3, 7, 8, 9.

Определим наиболее и наименее похожие между собой цифры. Для этого в паре цифр посчитаем сколько раз одна распознала как другая. В табл. 1 выписаны некоторые пары.

Таблица 1

Схожесть цифр при распознавании тестового набора

Пары	1-4	2-5	0-1	0-8	5-6	2-8	7-9	3-5	4-9	2-7
Ошибки	0	0	1	4	8	9	10	13	14	16

По таблице мы видим, что чаще всего сеть путает цифры 2 и 7, 4 и 9, а также, что сеть никогда не путает цифры 1 и 4, 2 и 5. Теперь становится понятным, почему на рис. 2 цифра 4 занимает 2 место.

Схожесть цифр мы можем определить следующим образом. Обучим сеть только на 2 цифрах, тогда мерой схожести станет ошибка разделения на классы. В силу не равного количества изображений каждой цифры, будем использовать относительную ошибку. Будем варьировать размером скрытого слоя и числом эпох обучения, для достижения наибольшей точности. Результаты расчетов представлены в табл. 2.

Таблица 2

Схожесть цифр при распознавании 2 классов

Пары	1-4	0-1	0-8	5-6	2-8	7-9	3-5	4-9	2-7
% ошибок	0.09	0.05	0.4	0.75	0.54	0.83	0.52	1	0.82

Данные табл. 1 и табл. 2 не совпадают, вследствие влияния остальных цифр. Тем не менее, они хорошо коррелируют между собой. Наименее похожими являются 0 и 1, что вполне отвечает интуитивному представлению. Человеку легко спутать 4 и 1, но для сети они наименее похожие. Интересно, что наиболее похожими становятся 4 и 9. Человек вряд ли спутал бы эти цифры. Данные таблиц 1 и 2 показывают, что в процессе распознавания сеть не идентифицирует отдельные элементы изображения, такие как закругления, хвостики и прямые линии. В связи с этим, точность распознавания в данной архитектуре ограничена. Без умения распознавать данные элементы, достигнутая точность в 98.25% кажется немислимой.

При распознавании 2 классов средняя ошибка распознавания составляет 0.55% (табл. 2), тогда как на полном наборе она достигает 1.75%. Проследим, как меняется ошибка при увеличении числа распознаваемых классов (рис. 3).

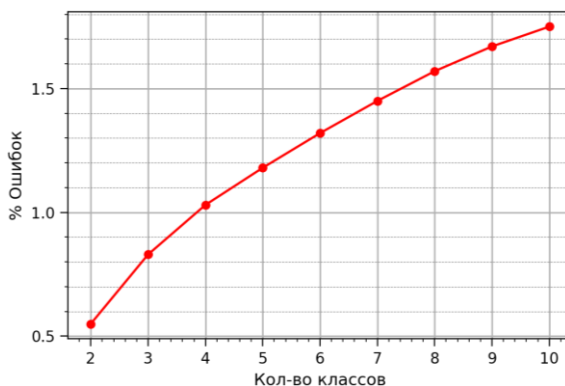


Рис. 3. Зависимость ошибки от числа распознаваемых классов.

Для достижения максимальной точности при увеличении количества распознаваемых классов необходимо увеличивать размер скрытого слоя. На рис. 4 представлен график изменения размера скрытого слоя от числа распознаваемых классов.

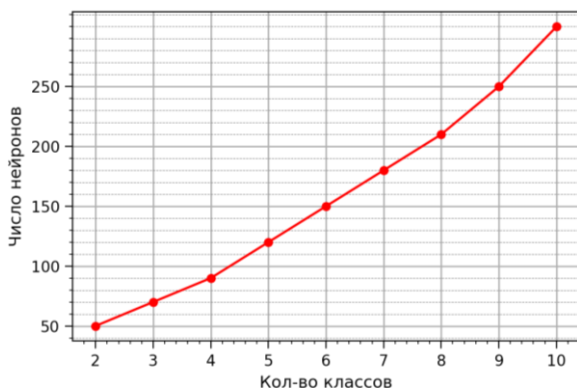


Рис. 4. Зависимость размера скрытого слоя от числа распознаваемых классов

Для разделения изображений на классы можно применять различные сети, содержащие меньше классов, чем требуется. Например, можно разделить все изображения цифр на 2 класса, потом еще на 2 и т.д. По графику 3 видно, что с увеличением числа классов прирост ошибки уменьшается. Данная тенденция позволяет сделать предположение о том, что нейронная сеть, содержащая много классов, покажет большую эффективность, чем последовательное применение нескольких сетей с меньшим числом классов. Это предположение подтверждается вычислительными экспериментами. Более того, расчеты показывают невозможность увеличить точность в спорной ситуации за счет 2х-классовых сетей.

Заключение

Мы изучили процесс распознавания рукописных цифр с помощью полносвязной двухслойной нейронной сети. Выяснили, что цифры 1 и 0 распознаются легко, а 2, 7 и 9 сложнее. В качестве неверного прогноза сеть чаще всего выдает 4 и 9.

Анализ схожести цифр показал нам, что нейронная сеть не оперирует такими понятиями как прямая линия, закругление, волнистая линия, и, тем не менее, достигает высокой точности в процессе распознавания.

Исследование разделения изображений по различным классам позволило оценить размер скрытого слоя необходимый для эффективного распознавания. Замедление роста ошибки при увеличении количества классов свидетельствует об эффективности сетей содержащих много классов.

Литература

1. Николенко С.И. Глубокое обучение / С.И. Николенко, А. Кадурын, Е. В. Архангельская. – СПб.: Питер, 2018. – 480 с.: ил. – (Серия «Библиотека программиста»).
2. Гудфеллоу Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль; пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил.
3. Рашид Т. Создаем нейронную сеть / Т. Рашид; пер. с англ. – СПб.: ООО «Альфа-книга», 2017. — 272 с.
4. Траск Э. Грокаем глубокое обучение / Э. Траск – СПб.: Питер, 2019. – 352 с.: ил. – (Серия «Библиотека программиста»)
5. Попов, М. И. Анализ многослойной полносвязной нейронной сети прямого распространения для распознавания рукописных цифр / М. И. Попов, Е. Н. Ковалева // Материалы LX отчетной научной конференции преподавателей и научных сотрудников ВГУИТ за 2021 год: В 3 частях, Воронеж, 08–09 февраля 2022 года / под ред. О.С. Корнеевой; Воронеж. гос. ун-т инж. технол. Том Часть 2. – Воронеж: Воронежский государственный университет инженерных технологий, 2022. – С. 79. – EDN СУВРСТ.