

**Национальный Исследовательский Университет ИТМО
Кафедра ВТ**

Лабораторная работа №4
Системы искусственного
интеллекта
Нейронные сети

Выполнил: Федоров Сергей
Группа: Р33113
Преподаватель: Болдырева Елена Александровна

Санкт-Петербург
2020 г.

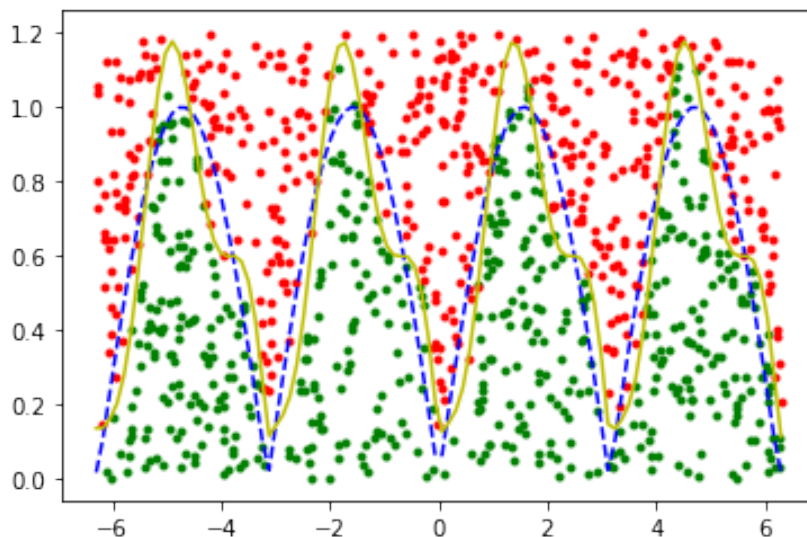
Задание: По предоставленным реализациям нейронных сетей, подобрать подходящие гиперпараметры для достижения лучшей оценки точности.

Вариант: (17 - номер в ису) % (4-1) = №3

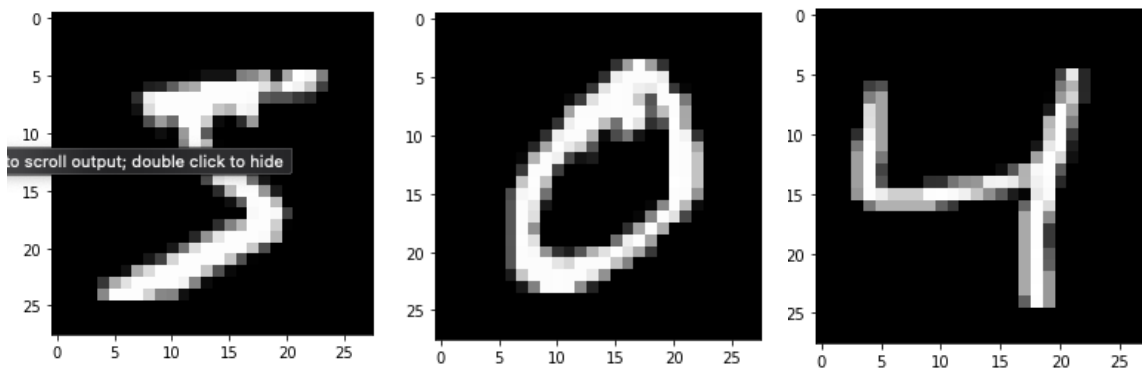
Absolute(Sin(x)) X: 6,3..6.3 Y: 0..1.2	Handwritten digits	Regularization L2, output layer activation type
--	--------------------	---

Данные:

По варианту в части **1** генерируются 2-мерные точки с и соответствующие заданной функции ответы (target).



По варианту в части **2** следует выбрать датасет **MNIST** и загрузить из любого доступного источника (я взял предоставляемый из пакета keras датасет с размерностью [70000, 784 (28x28)])



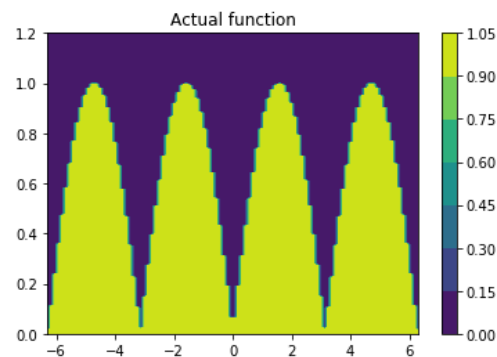
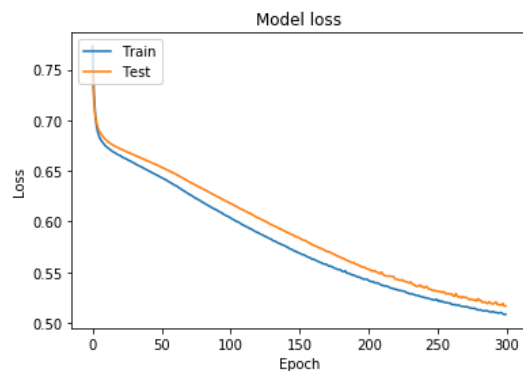
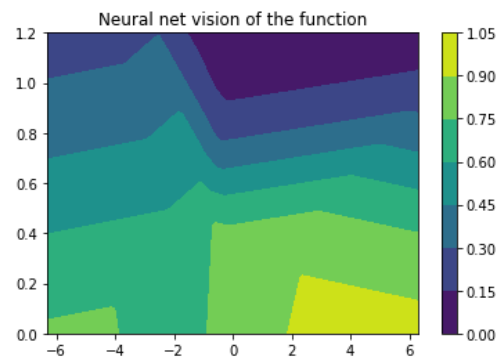
Выполнение:

Доступные параметры для подбора по варианту: Regularization L2, output layer activation
Другие параметры нам не доступны и это плохо, дальше мы поймем почему.

Часть 1:

Самый хороший и стабильный результат получался при выходном слое активации установленном в sigmoid (это не удивительно, ведь мы пытаемся предсказать принадлежность к классу). L2 reg = $1e-3$

Accuracy: 0.7699999809265137



Другие функции активации здесь выглядят ненужными и при таком большом шаге градиентного спуска все равно взрываются)))

Часть 2:

В данном случае мы уже работаем с данными гораздо большей размерности и более того проводим классификацию по множественным классам (в противоположность бинарной как в первой части).

Первое замечание, при данном кол-ве эпох и таком малом batch_size мы получаем достаточно сильный разброс по кривой схождения даже при одинаковых параметрах.

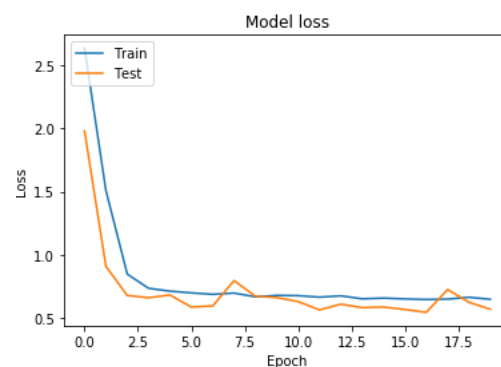
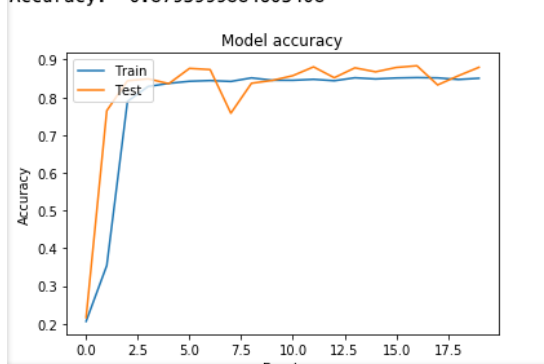
Второе замечание, при данных размерах сети (2 слоя) и отсутствия сверточных слоев, предсказание даже таких простых объектов как рукописные цифры - становится проблематичной задачей. Так сказать:



Учитывая характер поставленной задачи, стоит применить softmax как выходную активацию нашей сети. Однако при заданном оптимизаторе (обычный стохастический спуск) и с таким большим шагом как $1e-2$, процесс обучения не работает и сеть выдает одни и те же значения, предположу что просто в процессе обучения, буквально на первых шагах у нас взрывается градиент.

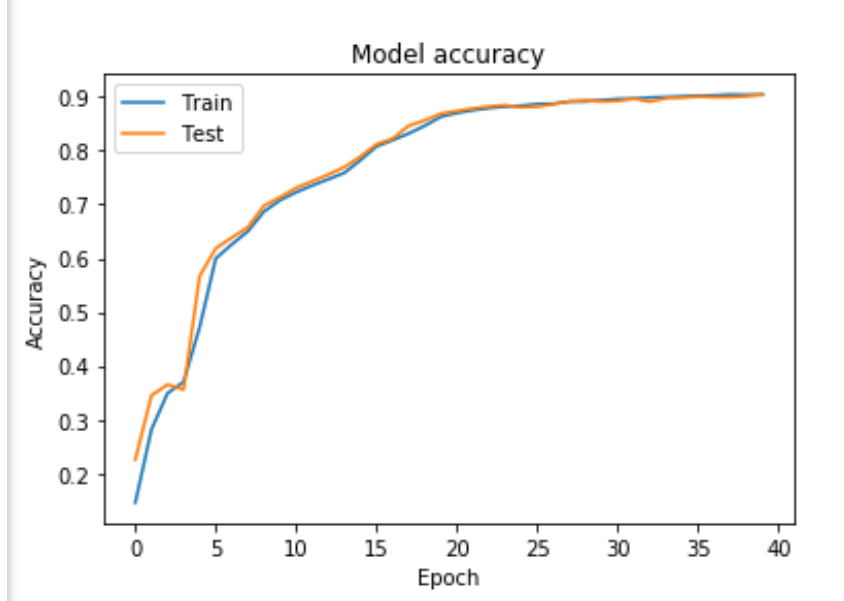
Лучшие значения полученные мною были достигнуты мною при $L2 \text{ reg} = 0.05$ и $\text{out_activation} = \text{sigmoid}$. Хотя, конечно, судя по графикам, это скорее случайный результат.

Accuracy: 0.8795999884605408



Выступив за пределы задания, поставил softmax и шаг градиента чуть поменьше:

Accuracy: 0.9039000272750854



Уверен что если еще поиграться со скрытыми слоями и оптимизаторами (например добавить момента), то можно будет достичь гораздо более впечатляющих результатов.

Вывод:

Помимо стандартных “сделал, понял” после лабораторной работы осталось несколько мыслей/выводов:

1. Мы были достаточно сильно ограничены во время выполнения работы.
2. Жалко что не дали попробовать другие типы нейронных сетей или не дали возможности (в разрезе лабораторной работы) написать сеть самостоятельно.
3. Обучение нейронной сети процесс затратный по ресурсам (как памяти так и времени) а так же сильно зависимый от изначально заданных гиперпараметров. Отчасти по этому в реальном мире распространена практика использования предобученных сетей.