

Лабораторная работа №5

Neural Networks+MNIST

Дисциплина: Системы Искусственного Интеллекта

Вариант 3

Выполнил: Чангалиди Антон

Группа: P33113

Преподаватель: Болдырева Е.А.

Цель лабораторной работы

Решить задачу многоклассовой классификации, используя в качестве тренировочного набора данных - набор данных MNIST, содержащий образы рукописных цифр.

Задание

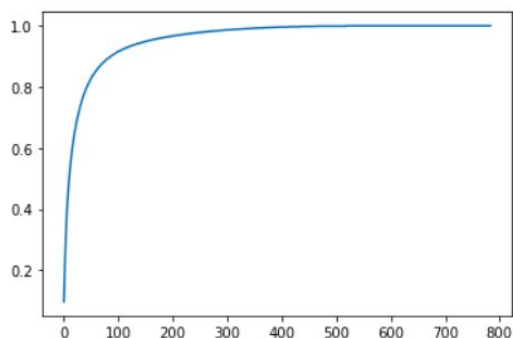
1. Используйте метод главных компонент для набора данных MNIST (train dataset объема 60000). Определите, какое минимальное количество главных компонент необходимо использовать, чтобы доля объясненной дисперсии превышала $0.80 + \text{номер_в_списке} \% 10$. Построить график зависимости доли объясненной дисперсии от количества используемых ГК
2. Введите количество верно классифицированных объектов класса номер_в_списке%9 для тестовых данных
3. Введите вероятность отнесения 5 любых изображений из тестового набора к назначенному классу
4. Определите Accuracy, Precision, Recall or F1 для обученной модели
5. Сделайте вывод про обученную модель

Порядок выполнения

1. Используйте метод главных компонент для набора данных MNIST (train dataset объема 60000). Определите, какое минимальное количество главных компонент необходимо использовать, чтобы доля объясненной дисперсии превышала $0.80 + \text{номер_в_списке} \% 10 = 0.88$. Построить график зависимости доли объясненной дисперсии от количества используемых ГК

Минимальное количество главных компонент, при которых доля объясненной дисперсии превышает 0.88 = 74

Out[6]: [



2. Введите количество верно классифицированных объектов класса номер_в_списке%9 для тестовых данных

Число верно классифицированных объектов класса 0 = 1359

3. Введите вероятность отнесения 5 любых изображений из тестового набора к назначенному классу

[12]:

	real_class	probability, that = 0
8013	8	0.064775
17801	7	0.192615
3983	2	0.073466
9229	9	0.021501
10774	5	0.076818

4. Определите Accuracy, Precision, Recall or F1 для обученной модели

```
Overall stat:  
Accuracy: 0.51  
Precision: 0.51  
Recall: 0.51  
F1: 0.51
```

```
By classes stat:
```

it[13]:

	Accuracy	Precision	Recall	F1
0	0.768	0.691	0.768	0.728
1	0.831	0.894	0.831	0.861
2	0.373	0.299	0.373	0.332
3	0.565	0.555	0.565	0.560
4	0.557	0.519	0.557	0.537
5	0.303	0.400	0.303	0.345
6	0.339	0.328	0.339	0.334
7	0.688	0.556	0.688	0.615
8	0.329	0.330	0.329	0.329
9	0.286	0.486	0.286	0.361

Код

https://github.com/TohaRhymes/ai_autumn_2020/tree/master/lab5_mnist_ensembles

Выводы

В ходе выполнения лабораторной была исследована применимость PCA и ансамблей для задач мульти-классификации. В конкретном примере использовался датасет рукописных цифр.

Текущая модель проявила не очень хорошее поведение на этом датасете, демонстрируя низкие характеристики (редко - выше 0.8). Самый высокий показатель 0 на единице (так как там всегда

прослеживается паттерн вертикальной черты, возможно, это берется во внимание (хотя, это не нейросети)).