

УНИВЕРСИТЕТ ИТМО

Факультет программной инженерии и компьютерной техники

Направление подготовки 09.03.04 Программная инженерия

Дисциплина «Системы искусственного интеллекта»

Лабораторная работа №4

Вариант 11

Студент

Кузнецов М. А.

P33131

Преподаватель

Авдюшина А. Е.

Санкт-Петербург, 2022 г.

Описание задания

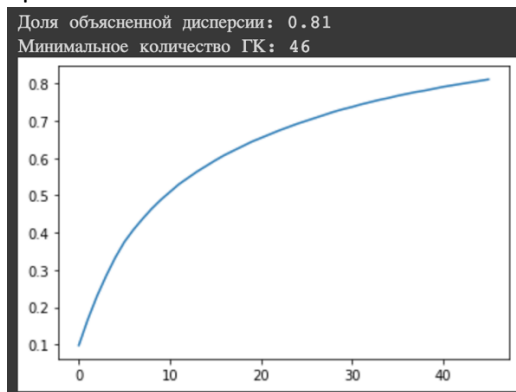
Цель: решить задачу многоклассовой классификации, используя в качестве тренировочного набора данных - набор данных MNIST, содержащий образы рукописных цифр.

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 $0.80 + \text{номер_в_списке} \% 10$.
2. Построить график зависимости доли объясненной дисперсии от количества используемых ГК.
3. Введите количество верно классифицированных объектов класса номер_в_списке%9 для тестовых данных.
4. Введите вероятность отнесения 5 любых изображений из тестового набора к назначенному классу.
5. Определите Accuracy, Precision, Recall or F1 для обученной модели.
6. Сделайте вывод про обученную модель.

Выполнение

Весь код доступен по ссылке: <https://shorturl.at/hBO29>

1. Необходимо минимум 46 главных компонент, чтобы доля объясненной дисперсии не превышала 0.81

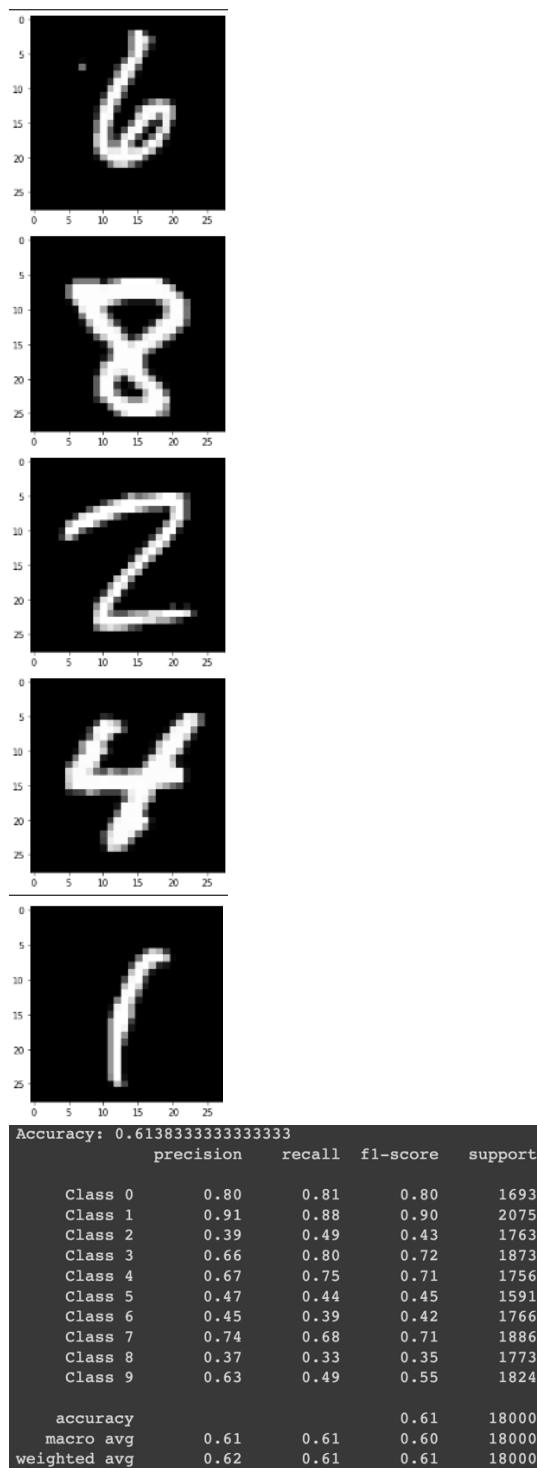


- 2.
3. Confusion_matrix (Класс 2)

```
[[1376  0 100  59  2  50  42  24  37  3]
 [  0 1836 32 10  8  8  50 121  8  2]
 [ 63 11 867 77 68 81 434 11 136 15]
 [ 23  6  65 1493 17 103 10 19 106 31]
 [ 13 12  87  7 1325 17 50 39 68 138]
 [ 91  9 175 213 55 699 11 38 267 33]
 [ 40 24 727 29 41 32 684 23 135 31]
 [ 55 57 22 18 59 34 167 1282 32 160]
 [ 46 46 134 286 64 433 43 28 589 104]
 [ 21 15 39 63 349 43 34 148 214 898]]
```

Количество верно классифицированных изображений: 867

- 4.
- ```
Predicted class: 2
True class: 6
Вероятность изображения No.7331 принадлежать к классу 2 -> 0.429
Predicted class: 8
True class: 8
Вероятность изображения No.8274 принадлежать к классу 8 -> 0.259
Predicted class: 5
True class: 2
Вероятность изображения No.4933 принадлежать к классу 5 -> 0.437
Predicted class: 4
True class: 4
Вероятность изображения No.3996 принадлежать к классу 4 -> 0.617
Predicted class: 1
True class: 1
Вероятность изображения No.7069 принадлежать к классу 1 -> 0.481
```



5.

## Выводы

В ходе выполнения данной лабораторной работы я:

- построил график зависимости для объясненной дисперсии
- реализовал многоклассовую классификацию образов рукописных цифр на алгоритме RandomForest.
- нашел параметры accuracy, precision и recall полученной модели