

федеральное государственное автономное образовательное учреждение высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

ОТЧЕТ

по лабораторной работе №5
Вариант №11

по дисциплине «**Системы искусственного интеллекта**»

Автор: Кулаков Н. В.

Факультет: ПИиКТ

Группа: Р3230

Преподаватель: Королёва Ю.А.



УНИВЕРСИТЕТ ИТМО

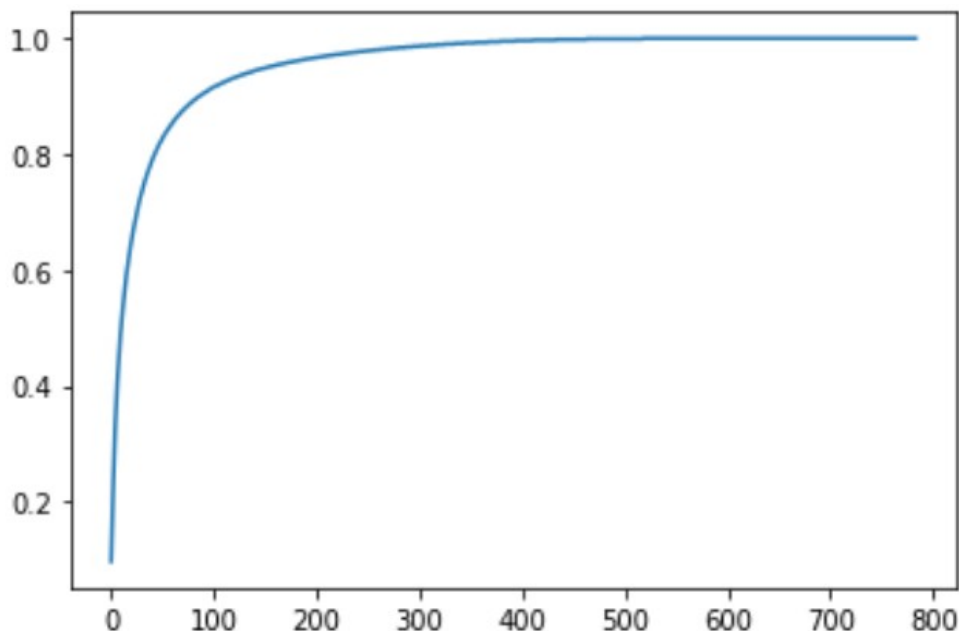
1) Текст задания.

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.

2) Выполнение.

График зависимости объясненной дисперсии:



Confusion matrix:

[603	0	7	15	4	72	21	9	9	3]
[3	824	16	3	2	6	6	1	11	3]
[11	6	587	26	19	15	18	7	39	5]
[7	5	30	658	3	33	6	5	39	8]
[4	4	17	0	592	2	16	18	15	100]
[80	10	14	94	24	302	37	26	67	50]
[20	5	30	7	17	23	653	1	8	8]
[12	15	12	0	21	5	1	665	13	59]
[10	15	25	78	15	50	9	8	468	24]
[10	9	3	25	234	29	7	144	37	288]]

```
# количество объектов класса 2 отнесенных ко 2 классу
```

```
# second 2 -> count of predicted 2s
```

```
cl = 11 % 9
```

```
print(cl)
```

```
m_confusion[cl][cl]
```

```
// 587
```

Вероятность отнесения 5 любых изображений)

```
from random import sample
indexes = sample(range(len(X_test)), k=5)
print(indexes)
test_objects = [X_test[i] for i in indexes]
for number, proba in zip(indexes, clf.predict_proba(test_objects)):
print(f"Image #{number}: class={y_pred[number]}, proba={round(max(proba), 4)}")
[7564, 2553, 5203, 7259, 3541]
Image #7564: class=6, proba=0.4753
Image #2553: class=8, proba=0.4711
Image #5203: class=0, proba=0.5694
Image #7259: class=9, proba=0.3346
Image #3541: class=0, proba=0.786
```

Accuracy, precision, recall, f1:

```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.79	0.81	0.80	743
1	0.92	0.94	0.93	875
2	0.79	0.80	0.80	733
3	0.73	0.83	0.77	794
4	0.64	0.77	0.70	768
5	0.56	0.43	0.49	704
6	0.84	0.85	0.84	772
7	0.75	0.83	0.79	803
8	0.66	0.67	0.66	702
9	0.53	0.37	0.43	786
accuracy			0.73	7680
macro avg	0.72	0.73	0.72	7680
weighted avg	0.73	0.73	0.73	7680

3) Вывод

В ходе выполнения данной лабораторной работы я построил график зависимости для объясненной дисперсии, а также реализовал многоклассовую классификацию с помощью метода опорных векторов. Кроме того, разобрался, как просматривать параметры модели.

4) Приложение

ссылка на репозиторий на гитхабе:

<https://github.com/zubrailx/University-ITMO/tree/main/Year-3/Artificial-intelligence/lab-5>