

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ
по лабораторной работе №8
по дисциплине «Машинное обучение»
Тема: Классификация (линейный дискриминантный анализ,
метод опорных векторов)

Студент гр. 6304

Иванов Д.В.

Преподаватель

Жангиров Т. Р.

Санкт-Петербург

2020

Цель работы

Ознакомиться с методами кластеризации модуля Sklearn.

Загрузка данных

1. Датасет скачан и загружен в датафрейм.

```
import pandas as pd
import numpy as np
data = pd.read_csv('iris.data', header=None)
```

2. Выделены данные и их метки, метки преобразованы к числам.

```
X = data.iloc[:, :4].to_numpy()
labels = data.iloc[:, 4].to_numpy()
le = preprocessing.LabelEncoder()
Y = le.fit_transform(labels)
```

3. Исходная выборка разбита на обучающую и тестовую.

```
X_train, X_test, y_train, y_test = train_test_split(X, Y,
test_size=0.5)
```

Линейный дискриминантный анализ

1. Проведена классификация наблюдений с использованием LDA.

Количество неправильно классифицированных наблюдений и точность классификации:

```
Wrong classified: 3
Score: 0.987
```

2. Построен график зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки.

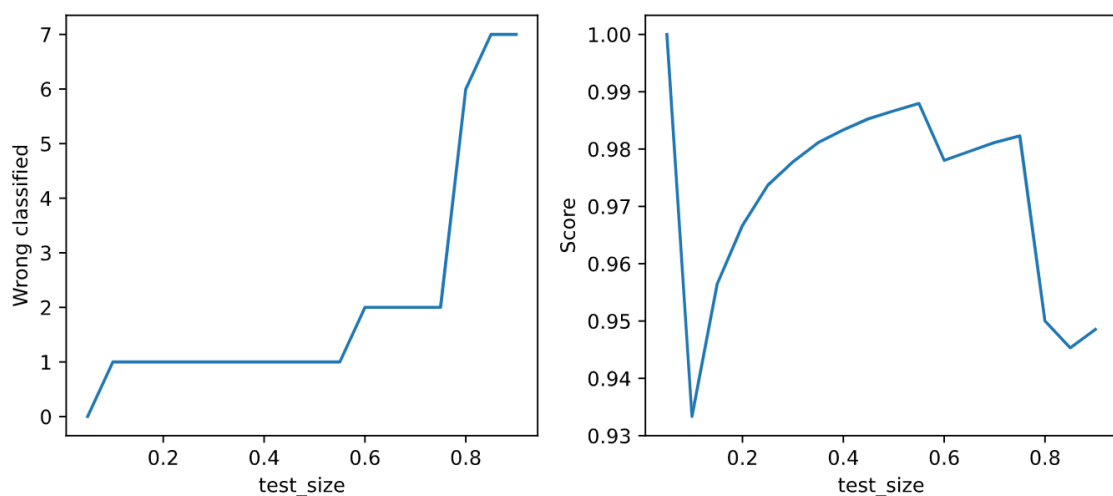


Рис. 1 — Графики зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки

3. Метод *transform* позволяет уменьшить размерность данных.

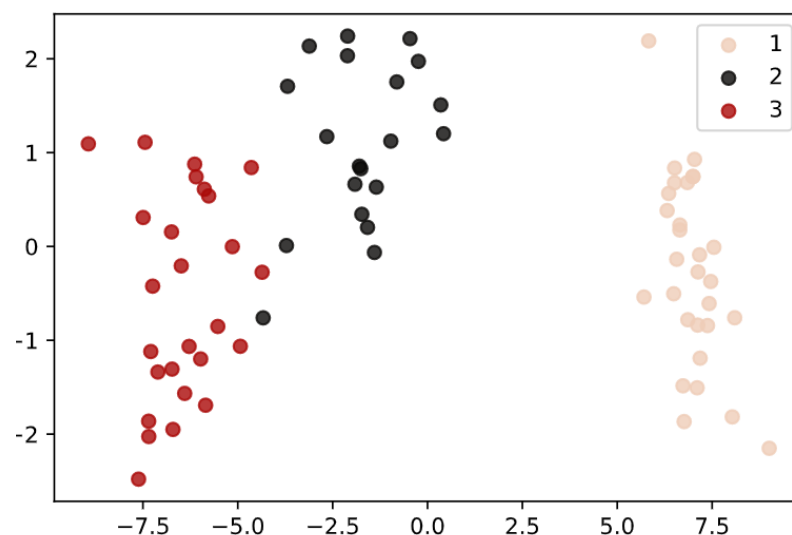


Рис. 2 — Результат работы метода *transform*

4. Исследована работа классификатора при различных параметрах *solver*, *shrinkage*

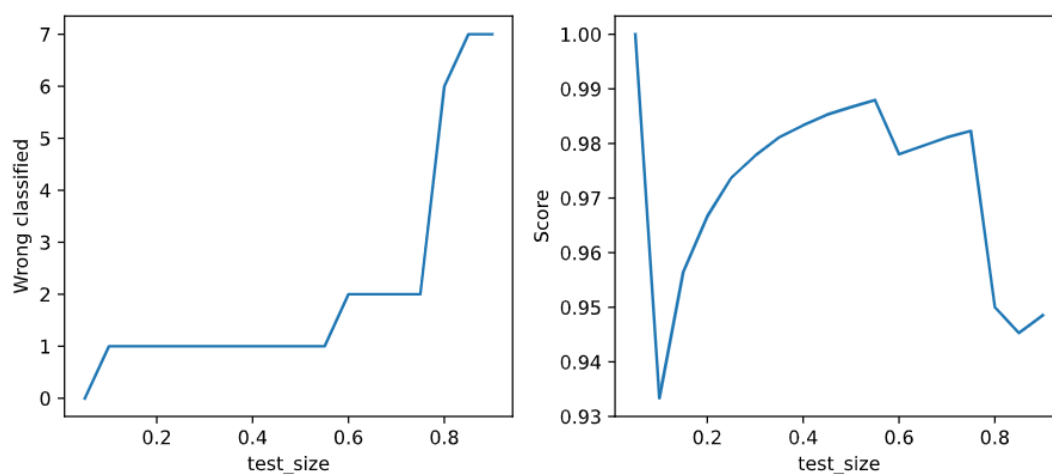


Рис. 3 — *solver=svd*, *shrinkage=no*

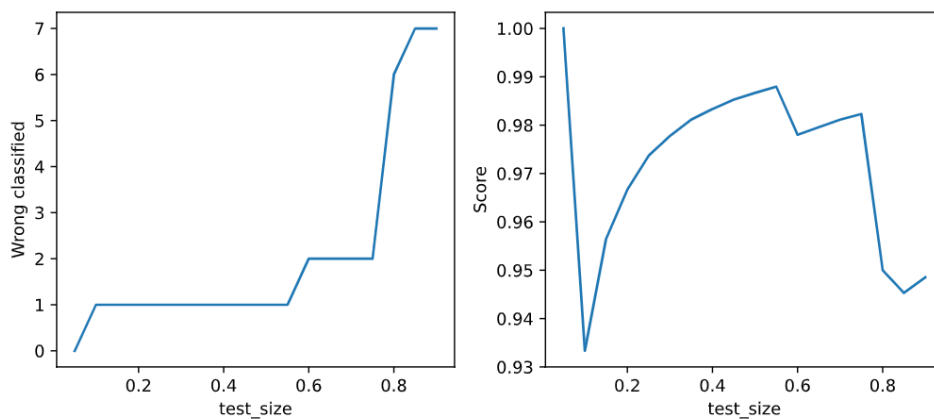


Рис. 4 — solver=lsqr, shrinkage=no

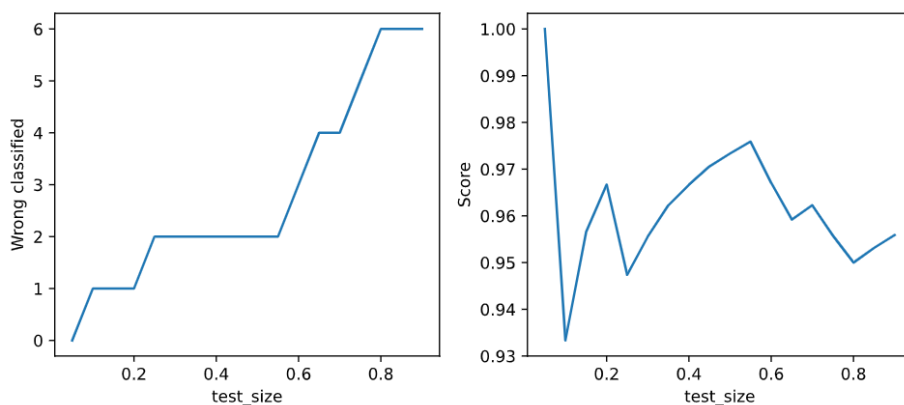


Рис. 5 — solver=lsqr, shrinkage=auto

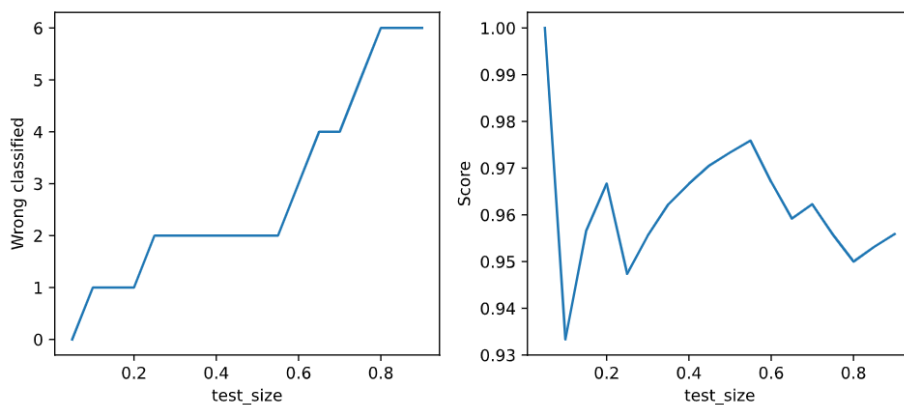


Рис. 6 — solver= eigen, shrinkage=auto

5. Классу с номером 1 задана априорная вероятность классу 0.7, остальным классам заданы равные априорные вероятности.

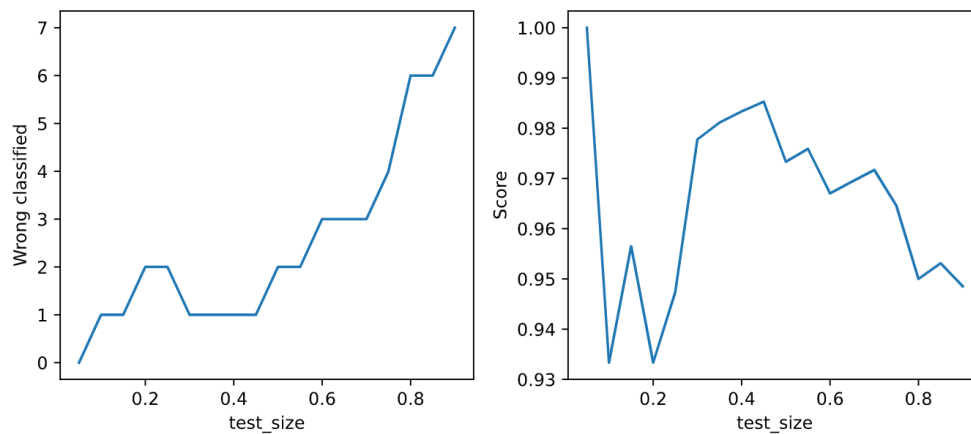


Рис. 7 — Графики зависимости от размера тестового набора

Метод опорных векторов

1. Проведена классификация SVM на тех же данных. Количество неверно классифицированных наблюдений, точность классификации, количество листьев дерева и глубина:

```
Wrong classified: 4
Score: 0.953
```

2. Построен график зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки.

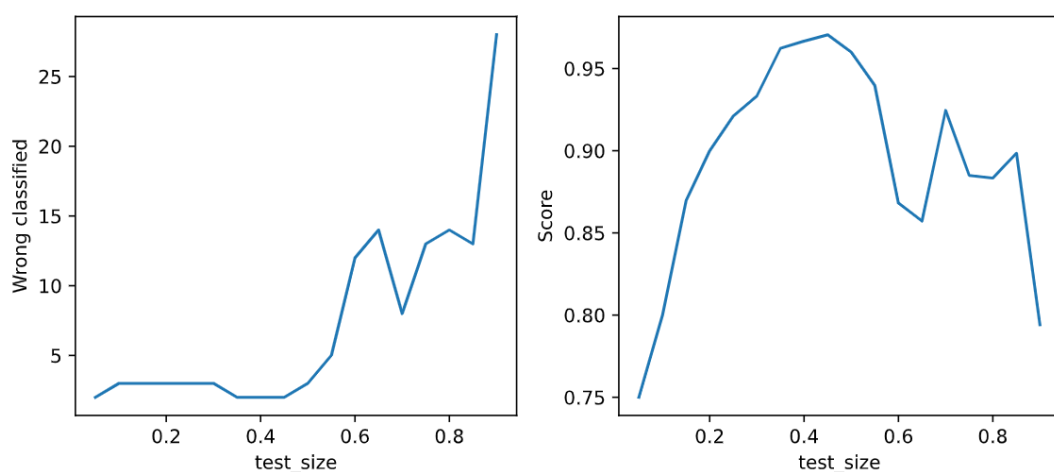


Рис. 8 — Графики зависимости от размера тестовой выборки

3. Исследована работа метода опорных векторов при различных значениях параметра.

а. *kernel* – тип ядра для использования внутри алгоритма. На исходных данных лучший результат показывает значение «linear».

<i>kernel</i>	Wrong classified	Score
linear	2	0.97
poly	6	0.95
rbf	4	0.95
sigmoid	54	0.33

б. *degree* – степень полиномиальной функции ядра (при *kernel*=poly).

<i>degree</i>	Wrong classified	Score
1	5	0.947
2	6	0.940
3	6	0.953
4	5	0.960
5	3	0.973
6	3	0.980
7	3	0.980

с. *max_iter* – ограничение на количество итераций.

<i>max_iter</i>	Wrong classified	Score
1	9	0.920
2	8	0.940

3	5	0.950
4	3	0.967
5	1	0.980
6	3	0.967
7	3	0.967
8	4	0.960
9	5	0.953
10	5	0.953

4. Исследована работа методов:

- a. NuSVC – имеет параметр для управления количеством опорных векторов.

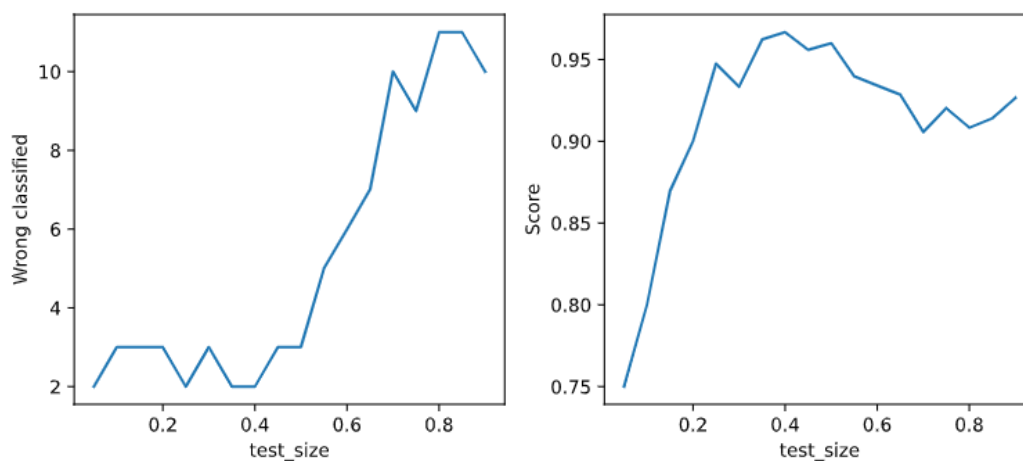


Рис. 9 — Графики зависимости от размера тестовой выборки NuSVC

- b. LinearSVC – аналогичен SVC с kernel=linear, но лучше масштабируется.

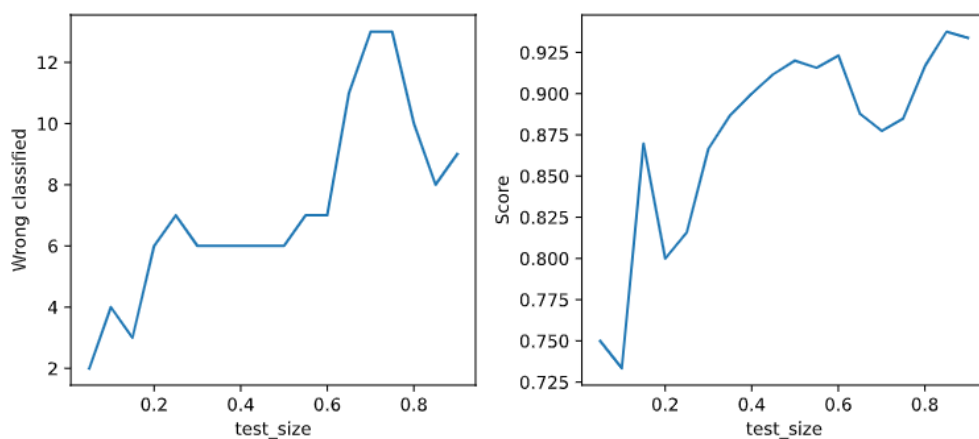


Рис. 9 — Графики зависимости от размера тестовой выборки LinearSVC

Вывод

В ходе выполнения лабораторной работы произведено знакомство с классификацией методом линейного дискриминантного анализа и методом опорных векторов модуля Sklearn.