

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

**ОТЧЕТ**  
**по лабораторной работе №8**  
**по дисциплине «Машинное обучение»**

Студенты гр. 6304

Преподаватель

\_\_\_\_\_  
\_\_\_\_\_

Тимофеев А.А.

Жангиров Т.Р.

Санкт-Петербург

2020

## Цель работы

Ознакомиться с методами классификации модуля Sklearn

## Ход работы

### Загрузка данных

1. Был создан датафрейм Pandas на основе загруженного датасета (<https://archive.ics.uci.edu/ml/datasets/iris>)
2. Были выделены данные и их метки, тексты меток были преобразованы к числам при помощи *LabelEncoder*.
3. Выборка была разбита на обучающую и тестовую при помощи *train\_test\_split*.

### Линейный дискриминантный анализ

1. Была проведена классификация данных методом *LinearDiscriminantAnalysis*, выведено количество неправильно классифицированных наблюдений (представлено на рисунке 1).

Wrong classified: 2

Рисунок 1 – Количество неправильно классифицированных наблюдений

2. Описание параметров и атрибутов метода *LinearDiscriminantAnalysis* представлено в таблицах 1 и 2 соответственно.

Таблица 1 – Описание параметров метода *LinearDiscriminantAnalysis*

Название	Описание	Принимаемые значения	Значение по умолчанию
solver	Метод решения	<i>svd, lsqr, eigen</i>	<i>svd</i>
shrinkage	Параметр усадки	<i>auto, float</i>	<i>None</i>
priors	Априорные вероятности классов	<i>array-like of shape (n_classes,)</i>	<i>None</i>

n_components	Число компонент для уменьшения размерности	<i>int</i>	<i>None</i>
store_covariance	Явно вычислить взвешенную ковариационную матрицу внутри класса при solver=svd	<i>Bool</i>	<i>False</i>
tol	Абсолютный порог для того, чтобы сингулярное число X считалось значимым	<i>float</i>	<i>1.0e-4</i>
covariance_estimator	Используется для оценки ковариационных матриц вместо того, чтобы полагаться на эмпирическую оценку ковариации	<i>covariance_estimator</i>	<i>None</i>

Таблица 2 – Описание атрибутов метода *LinearDiscriminantAnalysis*

Название	Описание	Тип возвращаемого значения
<i>coef_</i>	Вектор(ы) веса	<i>ndarray of shape (n_features,) or (n_classes, n_features)</i>
<i>intercept_</i>	Массив прерывания	<i>ndarray of shape (n_classes,)</i>

<i>covariance_</i>	Взвешенная матрица ковариаций внутри класса	<i>array-like of shape (n_features, n_features)</i>
<i>explained_variance_ratio_</i>	Процент дисперсии, объясняемый каждым из выбранных компонентов	<i>ndarray of shape (n_components,)</i>
<i>means_</i>	Средние по классам	<i>array-like of shape (n_classes, n_features)</i>
<i>priors_</i>	Априорные вероятности по классам	<i>array-like of shape (n_classes,)</i>
<i>scalings_</i>	Масштабирование признаков в пространстве, охватываемом центроидами классов.	<i>array-like of shape (rank, n_classes - 1)</i>
<i>xbar_</i>	Общее среднее	<i>array-like of shape (n_features,)</i>
<i>classes_</i>	Уникальные метки классов	<i>array-like of shape (n_classes,)</i>

3. Была определена точность классификации с помощью метода *score*, она составила 98%.
4. Были построены графики зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для метода *LinearDiscriminantAnalysis*. Графики представлен на рисунке 2.

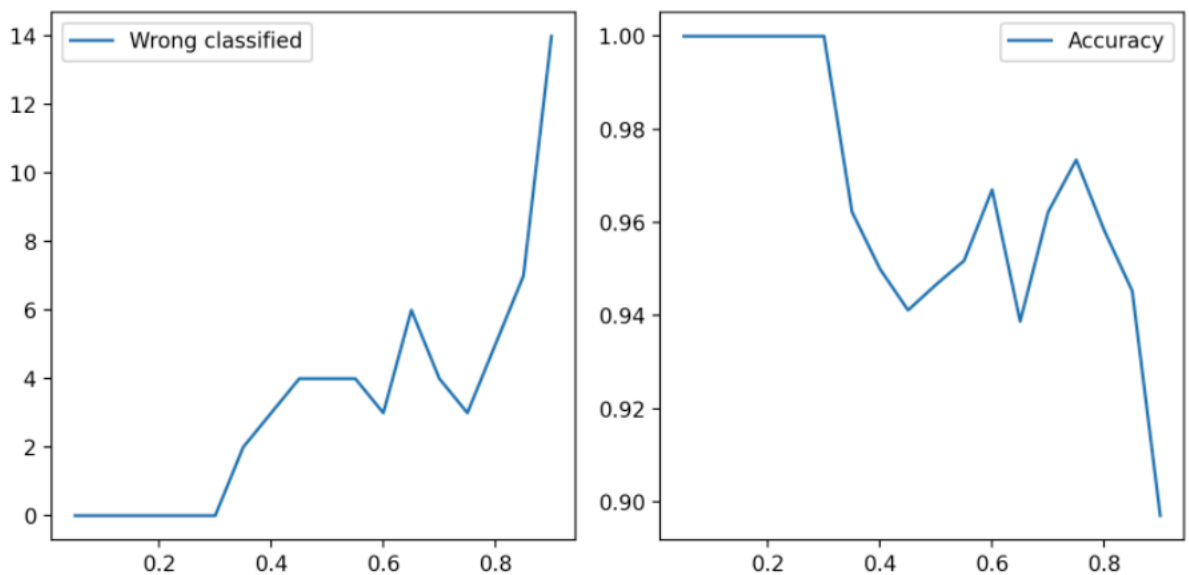


Рисунок 2 – Графики для метода *LinearDiscriminantAnalysis*

Точность классификации не падает с увеличением тестовой выборки. Скорее всего такая хорошая классифицируемость данных связана с их распределением в выборке.

- Метод *transform* используется для проецирования данных с целью максимизации дисперсии. График полученной проекции представлен на рисунке 3

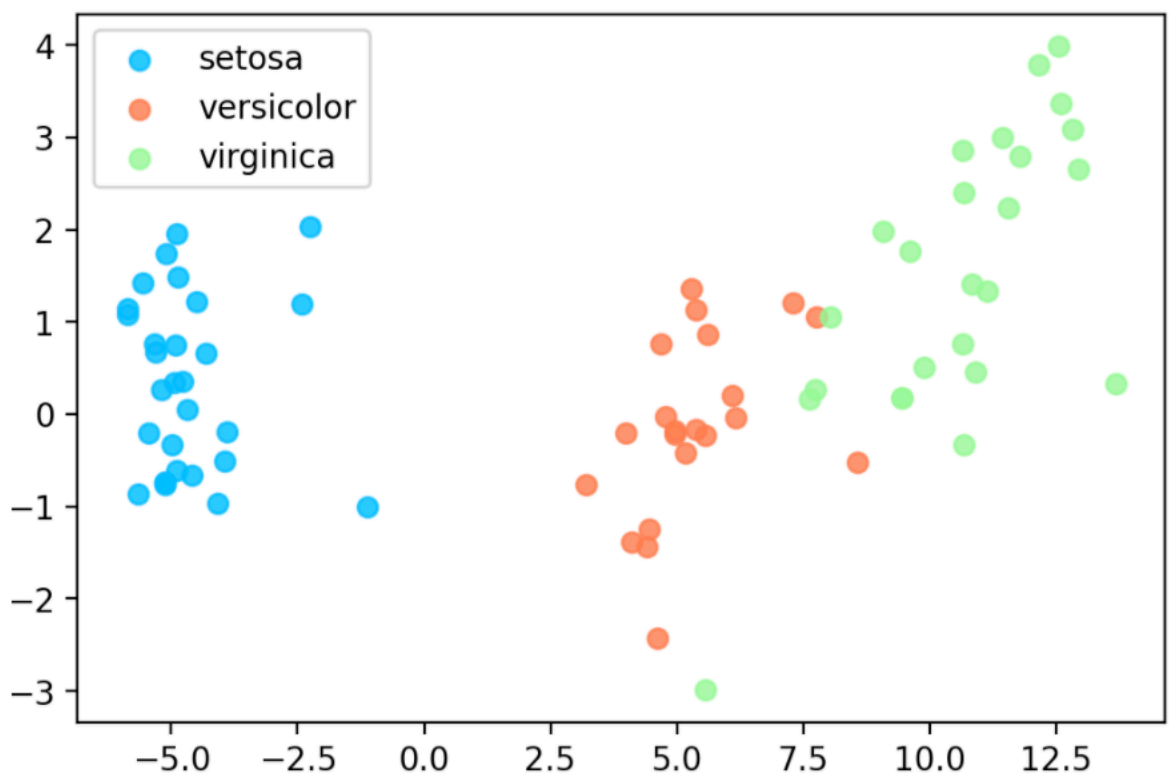


Рисунок 3 – Результат метода *transform*

6. Было проведено исследование классификации с различными параметрами solver. Результаты представлены на рисунках 4 – 6.

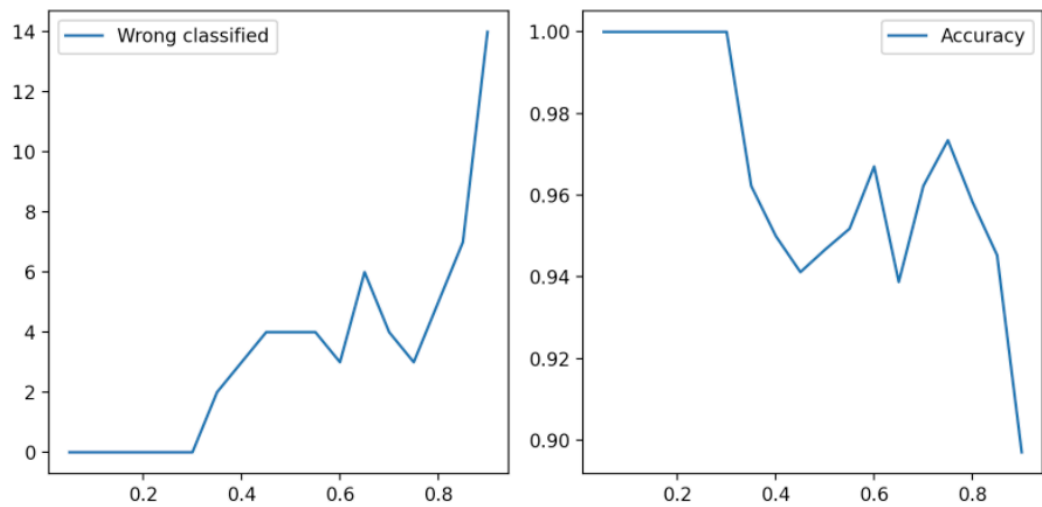


Рисунок 4 – Результаты с solver = svd

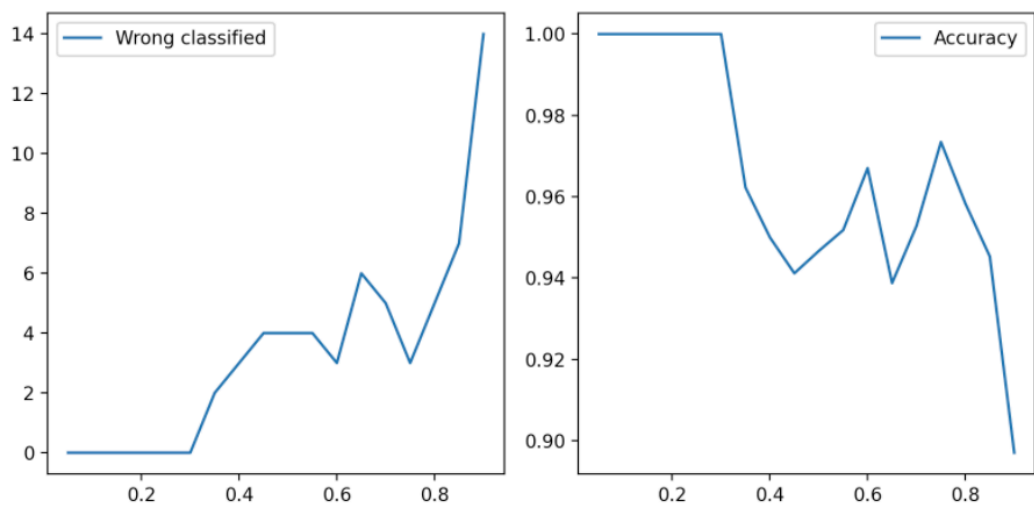


Рисунок 5 – Результаты с solver = lsqr

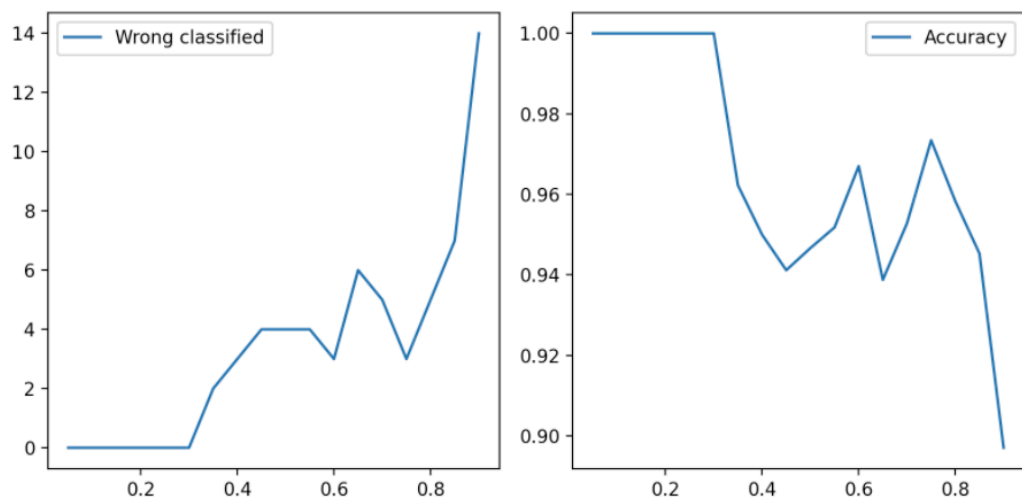


Рисунок 6 – Результаты с solver = eigen

Как видно из графиков классификация с разными параметрами solver дает практически идентичные результаты.

7. Эксперименты с параметром усадки показали небольшое ухудшение результатов классификации с его увеличением.
8. Была проведена классификация с вручную заданными априорными вероятностями классов. Результат представлен на рисунке 7.

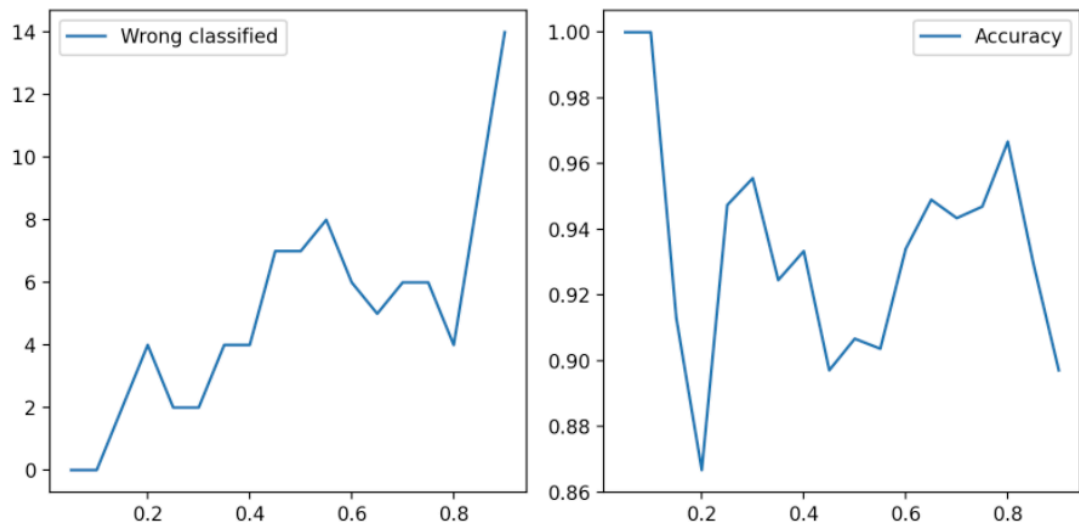


Рисунок 7 – Результаты классификации

Результат идентичен результату без указания вероятностей.

### Метод опорных векторов

1. Была выполнена классификация методом *SVC* на тех же данных. Точность составила 95%.
2. Были выведены опорные вектора, их индексы и количество опорных векторов для каждого класса.
3. Были построены графики зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для метода *SVC*. Графики представлен на рисунке 8.

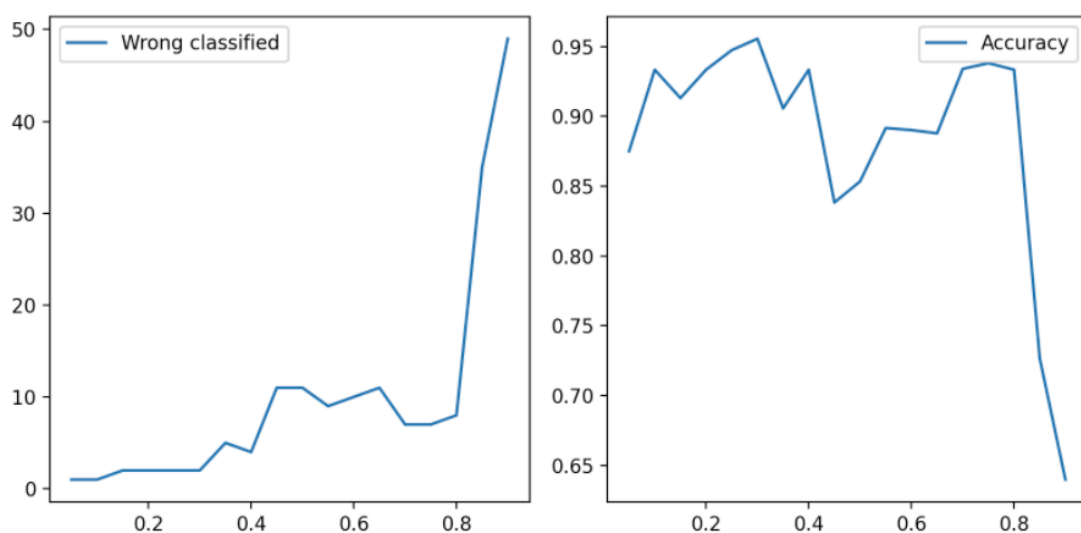


Рисунок 8 – Графики для метода *SVC*

4. Было проведено исследование работы метода с разными значениями параметра *kernel*. Лучшие результаты для каждого ядра представлены в таблице 3.

Таблица 3 – Метод *SVC* с разными ядрами

Ядро	Неправильно классифиц.	Точность
<i>linear</i>	0	100%
<i>poly</i>	2	98%
<i>rbf</i>	2	95%
<i>sigmoid</i>	76	33%

5. Было проведено исследование работы метода с разными значениями параметра *degree*. Лучшие результаты для каждого значения представлены в таблице 3.

Таблица 4 – Метод *SVC* с разными степенями функции ядра

Степень	Неправильно классифиц.	Точность
1	2	96%
2	0	100%
3	2	98%



4	2	96%
---	---	-----

6. Было проведено исследование работы метода с разными значениями параметра *max\_iter*. Лучшие результаты для каждого значения представлены в таблице 3.

Таблица 4 – Метод *SVC* с разным ограничением итераций

Ограничение	Неправильно классифиц.	Точность
1	4	97%
3	3	98%
5	2	98%
9	0	100%

7. Было проведено сравнение методов *SVC*, *NuSVC*, *LinearSVC*. Лучшие результаты для каждого метода представлены в таблице 3.

Таблица 4 – Метод *SVC* с разным ограничением итераций

Метод	Неправильно классифиц.	Точность
<i>SVC</i>	2	96%
<i>NuSVC</i>	5	95%
<i>LinearSVC</i>	5	96%

Метод *LinearSVC* идентичен методу *SVC* с линейным ядром. Метод *NuSVC* в отличие от метода *SVC* имеет параметр, отвечающий за количество опорных векторов, а также использует в себе другие математические вычисления.

## Выводы

В ходе выполнения данной лабораторной работы было произведено знакомство с методами классификации модуля Sklearn. Классификация

производилась с помощью методов *LinearDiscriminantAnalysis*, *SVC*, *NuSVC* и *LinearSVC*.