**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МОЭВМ**

отчет

**по лабораторной работе №8**

**по дисциплине «Машинное обучение»**

Тема: **Классификация (линейный дискриминантный анализ, метод опорных векторов)**

|  |  |  |
| --- | --- | --- |
| Студент гр. 8304 |  | Холковский К.В |
| Преподаватель |  | Жангиров Т. Р. |

Санкт-Петербург

2021

**Цель работы**

Ознакомиться с методами классификации модуля Sklearn

**Ход работы**

**Линейный дискриминантный анализ**

1. Провели классификацию используя LDA



Рис 1 – Результаты классификации

Таблица 1 - Параметры

|  |  |
| --- | --- |
| Параметр | Смысл |
| **solver** | Используемый решатель |
| **shrinkage** | Параметр усадки |
| **priors** | Класс априорных вероятностей. |
| **n\_components** | Количество компонентов для уменьшения размерности. |
| **store\_covariance** | Если True, явно вычислить взвешенную ковариационную матрицу внутри класса, когда решатель - «svd». Матрица всегда вычисляется и сохраняется для других решателей. |
| **tol** | Абсолютный порог для того, чтобы единичное значение X считалось значимым, используется для оценки ранга X. Измерения, единичные значения которых не значимы, отбрасываются. Используется, только если решатель - «svd». |
| **covariance\_estimator** | Используется для оценки ковариационных матриц вместо того, чтобы полагаться на эмпирическую оценку ковариации. |

Таблица 2 - Атрибуты

|  |  |
| --- | --- |
| Атрибуты | Смысл |
| **coef\_** | Вектор (ы) веса |
| **intercept\_** | Срок перехвата. |
| **covariance\_** | Взвешенная внутриклассовая ковариационная матрица. |
| **explained\_variance\_ratio\_** | Процент отклонения, объясняемый каждым из выбранных компонентов. |
| **means\_** | Классовые средние. |
| **priors\_** | Приоры класса |
| **scalings\_** | Масштабирование объектов в пространстве, охватываемом центроидами классов. Доступно только для решателей «svd» и «eigen». |
| **xbar\_** | Общее среднее. Присутствует, только если решатель - «svd». |
| **classes\_** | Уникальные метки классов. |
| **n\_features\_in\_** | Количество деталей, видимых во время посадки. |
| **feature\_names\_in\_** | Названия особенностей, замеченных во время посадки. Определяется только тогда, когда X имеет имена функций, которые являются строками. |

1. Построили график зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LDA.

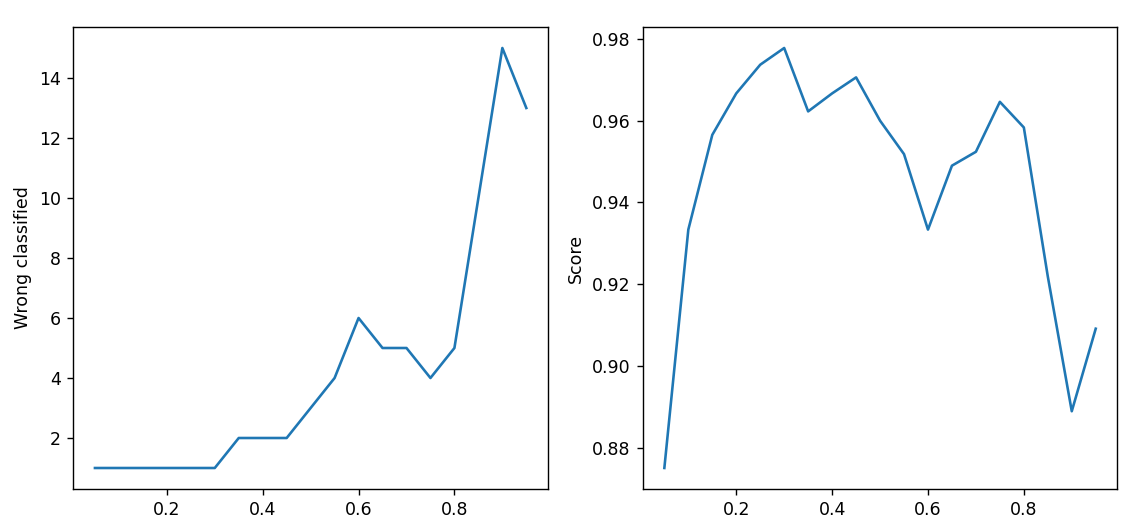


Рис 2 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LDA

1. Трансформировали данные

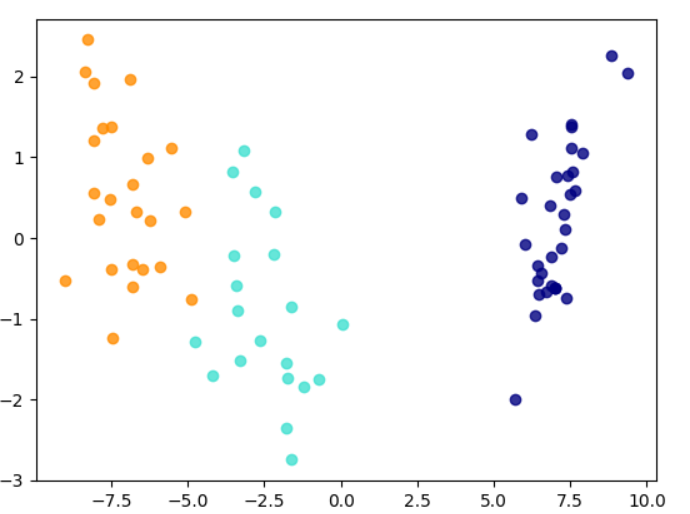


Рис 3 – Результат трансформации

1. Исследовали работу при различных параметрах solver, shrinkage.

Solver:

* ‘svd’: разложение по сингулярным значениям (по умолчанию). Не вычисляет ковариационную матрицу, поэтому этот решатель рекомендуется для данных с большим количеством функций.
* ‘lsqr’: решение по методу наименьших квадратов. Можно комбинировать с оценкой усадки или настраиваемой ковариационной оценкой.
* ‘eigen’: разложение на собственные значения. Можно комбинировать с оценкой усадки или настраиваемой ковариационной оценкой.

Shrinkage:

* none: no shrinkage (default).
* ‘auto’: automatic shrinkage using the Ledoit-Wolf lemma.
* float between 0 and 1: fixed shrinkage parameter.

1. Построили график зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LDA c параметром priors = [0.2, 0.7, 0.2].

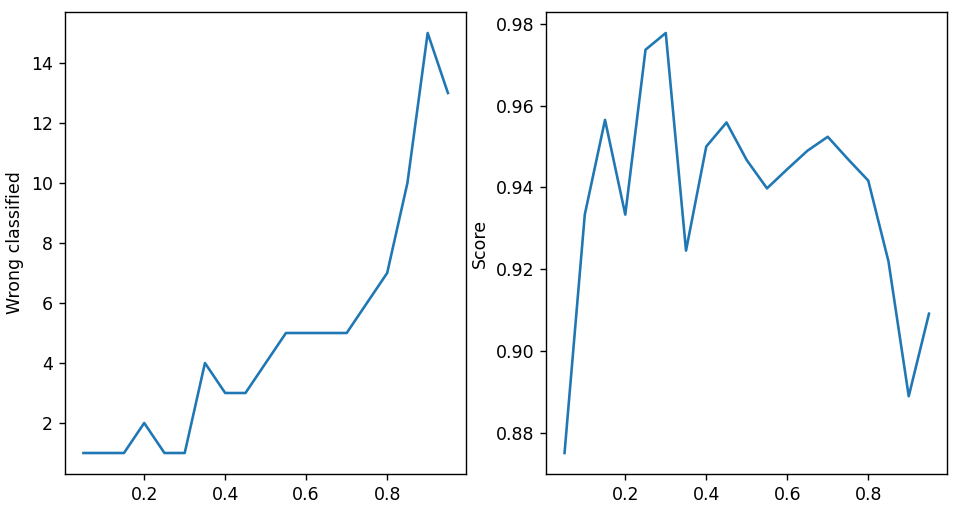


Рис 4 **-** График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LDA c параметром priors = [0.2, 0.7, 0.2]

**Метод опорных векторов**

1. Провели классификацию методом SVM

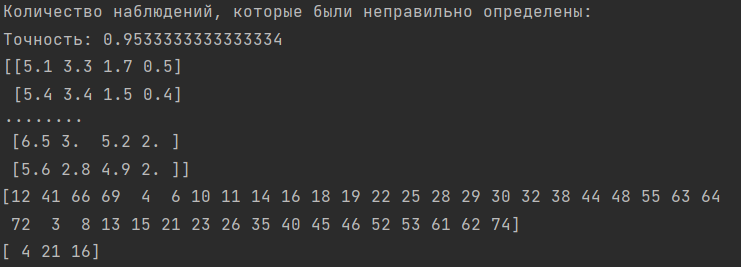


Рис 5 – Результаты классификации

1. Построили график зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для SVС

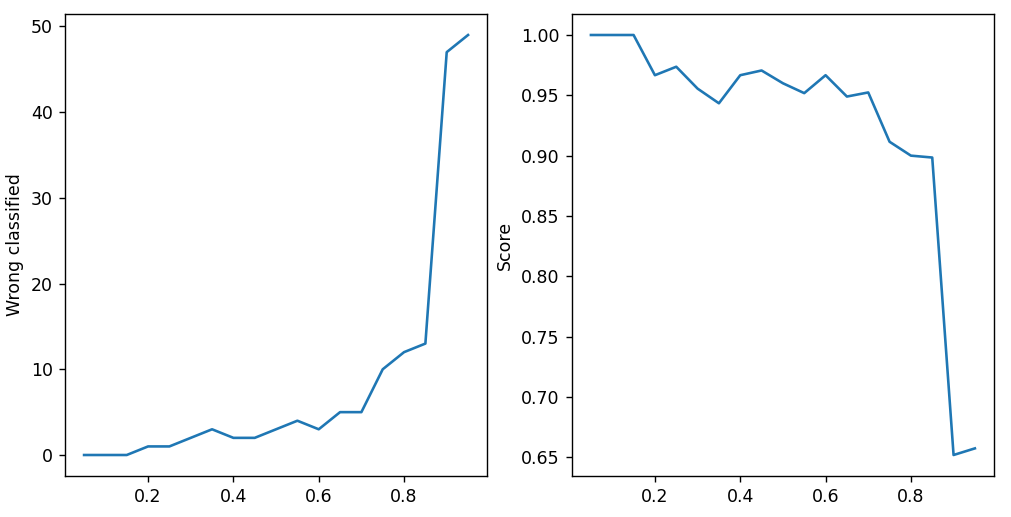


Рис 6 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для SVС

1. Исследовали работу SVM при различных значениях kernel, degree, max\_iter

kernel :

* 'linear'
* 'poly'
* 'rbf'
* 'sigmoid'
* 'precomputed'

degree: degree of the polynomial kernel function

max\_iter: hard limit on iterations within solver

1. Провели исследовали для методов NuSVC и LinearSVC

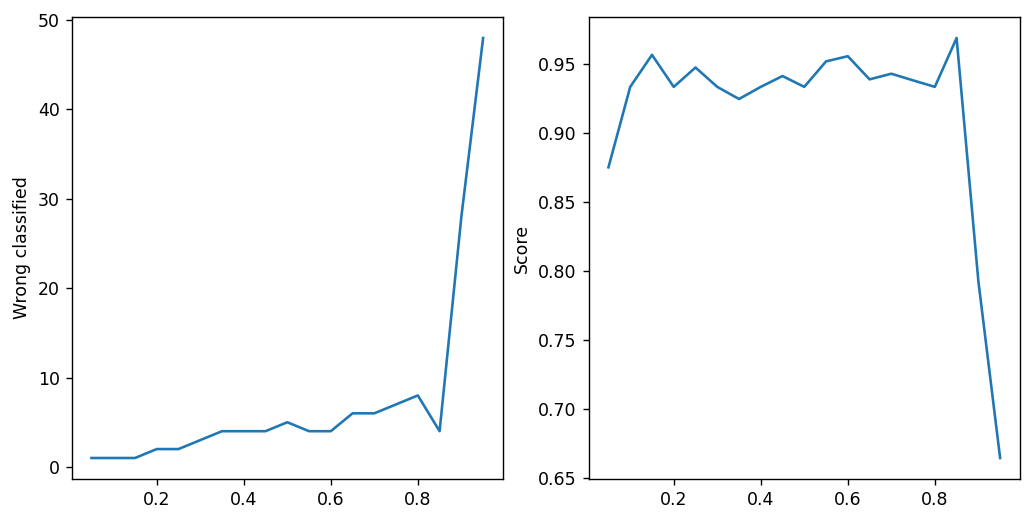


Рис 7 **-** График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LinearSVС

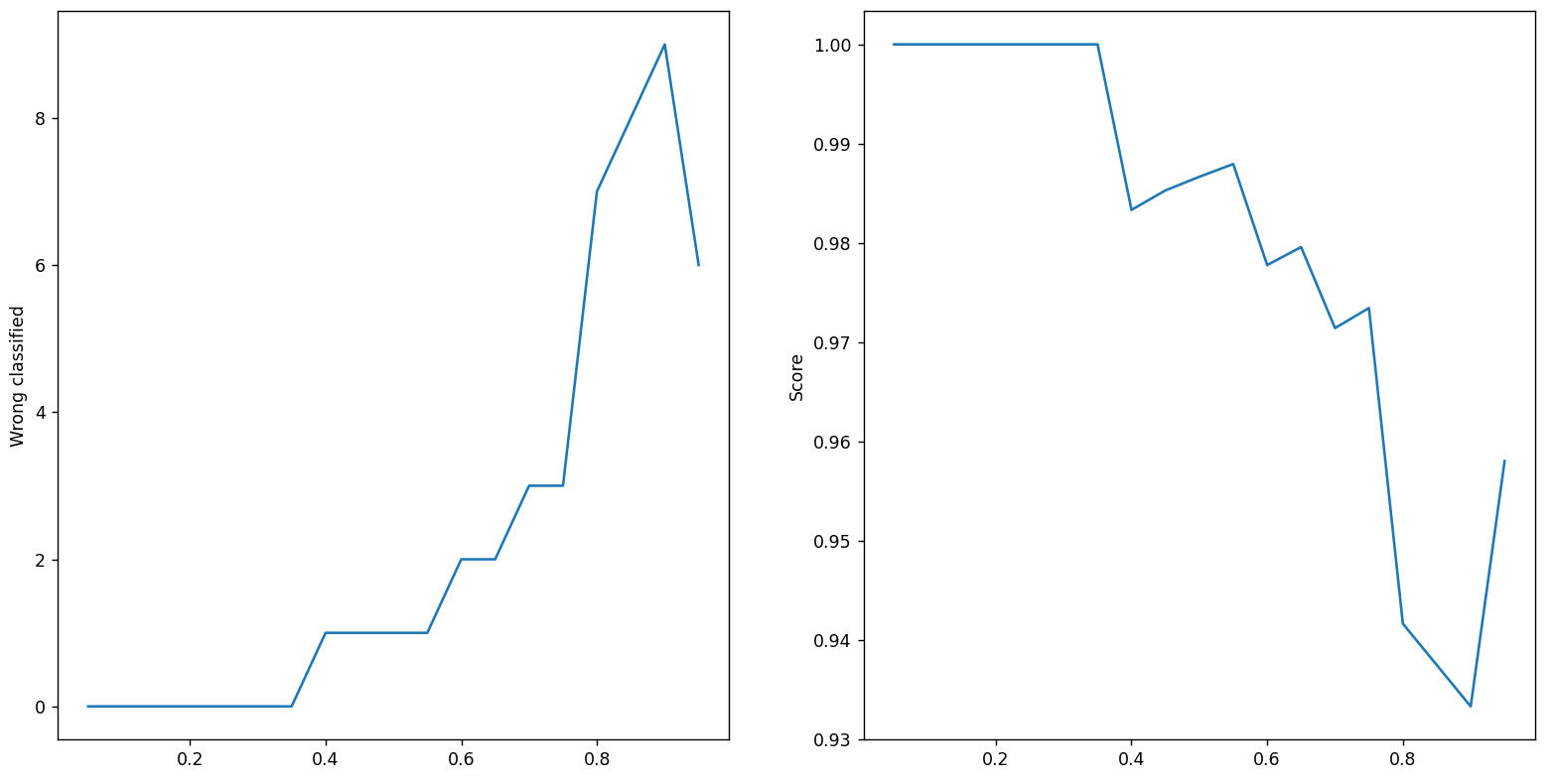


Рис 8 **-** График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для NuSVС

**Вывод**

Ознакомились с методами классификации модуля Sklearn.