

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

```
Количество наблюдений, которые были неправильно определены: 3  
Точность: 0.9866666666666667
```

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.

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

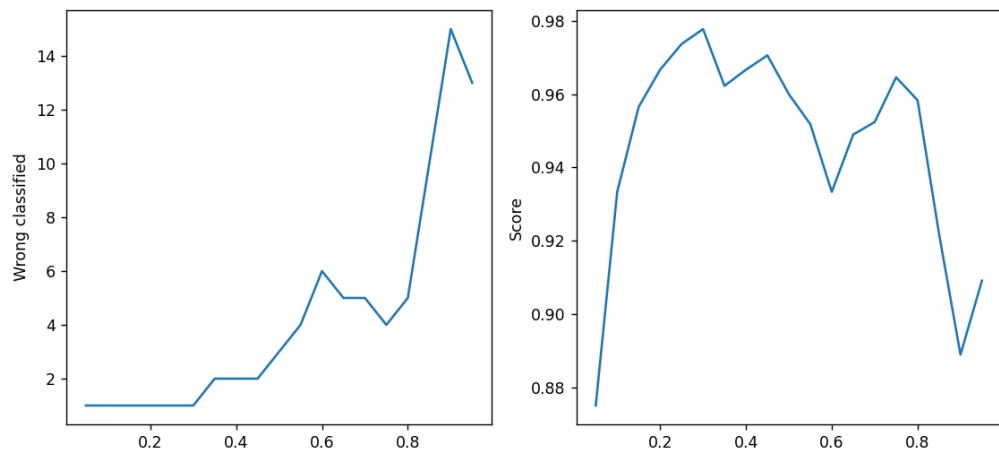


Рис 2 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LDA

3. Трансформировали данные

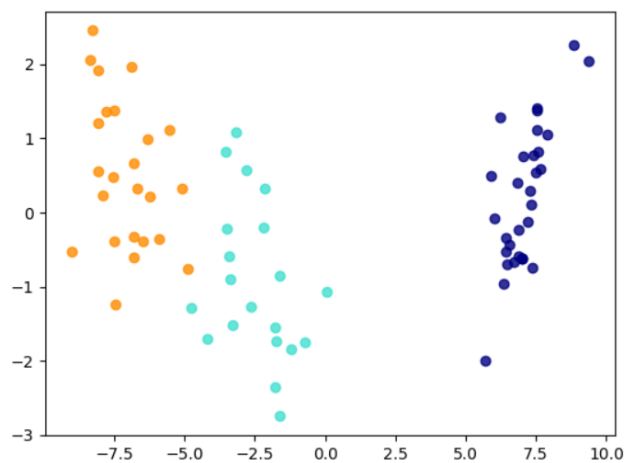


Рис 3 – Результат трансформации

4. 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.

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

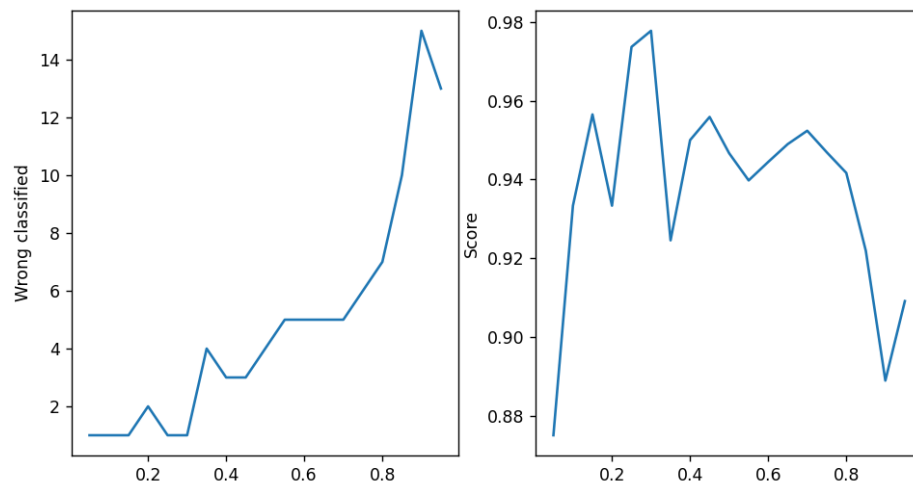


Рис 4 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LDA с параметром priors = [0.2, 0.7, 0.2]

Метод опорных векторов

1. Провели классификацию методом SVM

```

Количество наблюдений, которые были неправильно определены:
Точность: 0.9533333333333334
[[5.1 3.3 1.7 0.5]
 [5.4 3.4 1.5 0.4]
 .....
 [6.5 3.  5.2 2. ]
 [5.6 2.8 4.9 2. ]]
[12 41 66 69  4  6 10 11 14 16 18 19 22 25 28 29 30 32 38 44 48 55 63 64
 72  3  8 13 15 21 23 26 35 40 45 46 52 53 61 62 74]
[ 4 21 16]

```

Рис 5 – Результаты классификации

2. Построили график зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для SVC

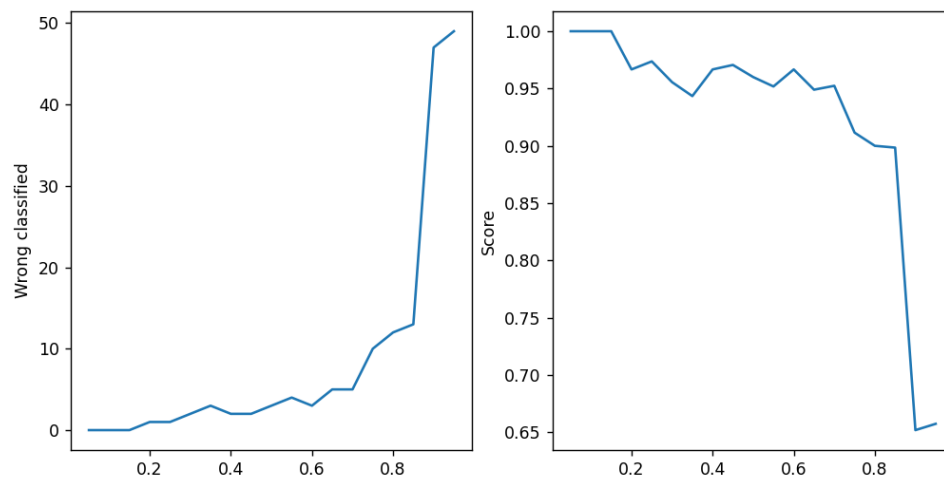


Рис 6 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для SVC

3. 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

4. Провели исследование для методов NuSVC и LinearSVC

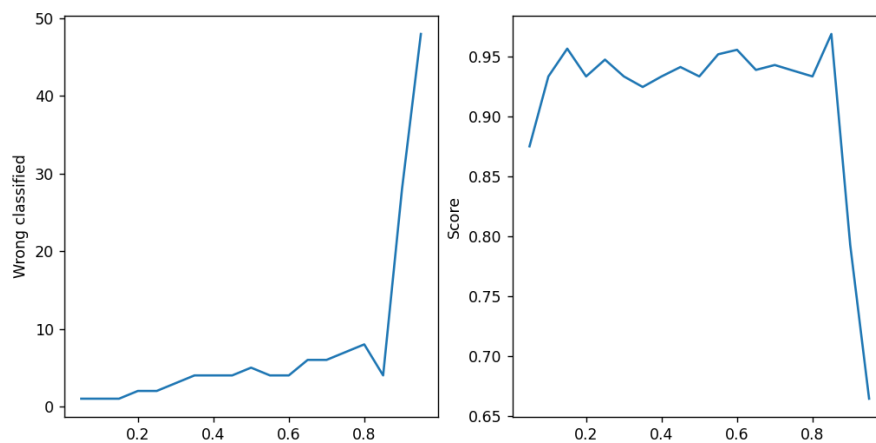


Рис 7 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для LinearSVC

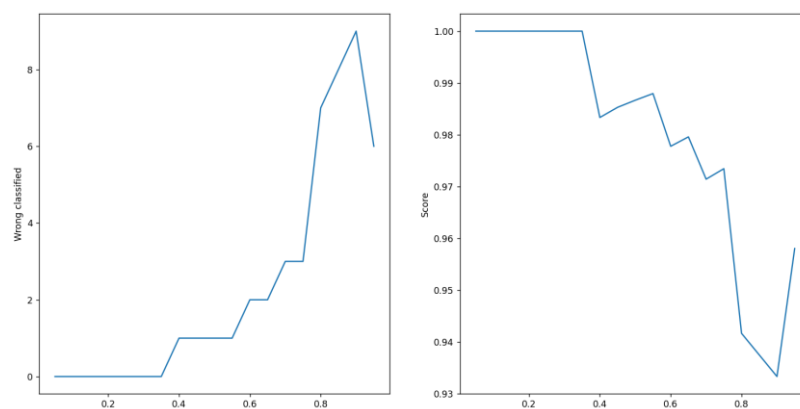


Рис 8 - График зависимости неправильно классифицированных наблюдений и точности классификации от размера тестовой выборки для NuSVC

Вывод

Ознакомились с методами классификации модуля Sklearn.