

درس شناسایی الگو

گزارش پروژه شماره 2

دسته بندی داده ها با استفاده از روشهای مختلف

نام و نام خانوادگی دانشجو:

فاطمه پارسا

بهمن 1399

مقدمه:

پروژه دوم درس پردازش الگو مربوط به بررسی کارآیی دسته بندی های مختلف برای هدف دسته بندی داده ها است. در این پروژه به ترتیب دسته بندی های LDA، Bayesian، PCA، SVM و min_cut بررسی می شود.

قسمت اول :

150 داده در سه دسته متفاوت با نام های (Iris_setosa, Iris_versicolor, Iris_virginica) دسته بندی شده اند و شامل چهار ویژگی (طول کاسبرگ، عرض کاسبرگ، طول گلبرگ و عرض گلبرگ است).

الف:

در قسمت اول با استفاده از دسته بندی LDA داده ها را دسته بندی می کنیم. LDA به این صورت عمل می کند که بعد داده ها را با توجه به میانگین کلاس ها کاهش می دهد. LDA روی تعداد بعد های جدید محدودیت دارد و بعد جدید باید حداقل یک واحد کمتر از تعداد دسته ها باشد. در این مثال تعداد بعد باید حداکثر 2 باشد.

ابتدا دیتاهای test را جدا کرده و با روش k_fold (K=2) دیتاهای باقی مانده را به دو قسمت train و validation تقسیم کرده ایم. درصد دیتای تست را با مقدار های 10٪ و 20٪ و 30٪ جدا کرده و هر بار با استفاده از LDA ابعاد را به 2 بعد کاهش داده و کلاس ها را با رنگ های متفاوت بر روی نمودار پراکنندگی نمایش داده ایم.

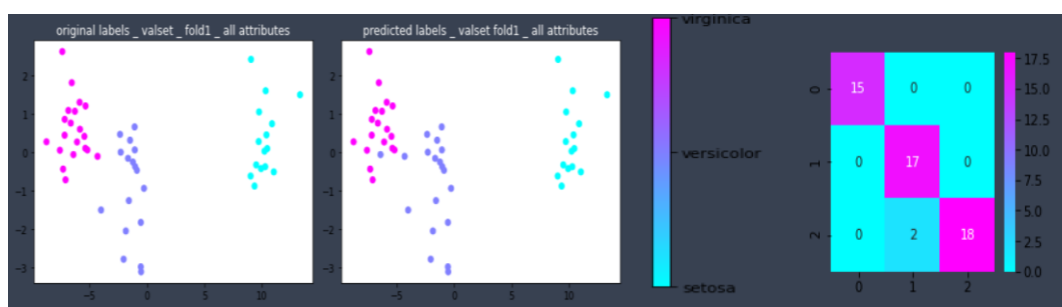
بار دیگر دسته بندی را با کاهش تعداد ویژگی ها انجام دادیم و نتایج به صورت زیر است:

Data=validation set

test_size=0.3

fold=1

accuracy=96%

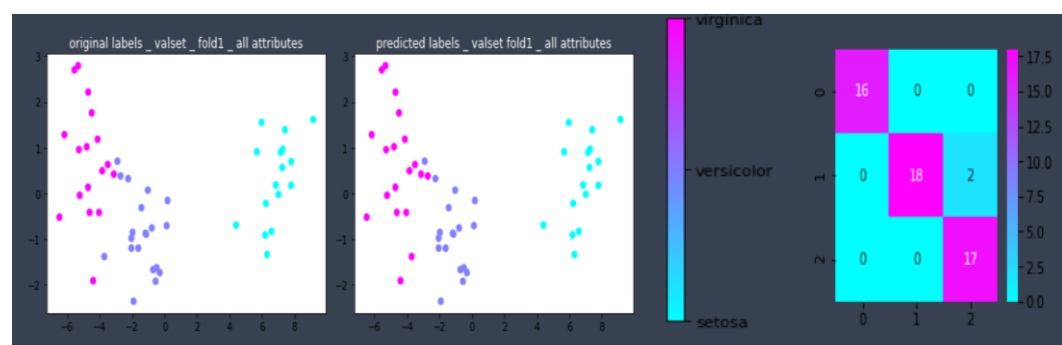


Data=validation set

test_size=0.3

fold=2

accuracy =96%



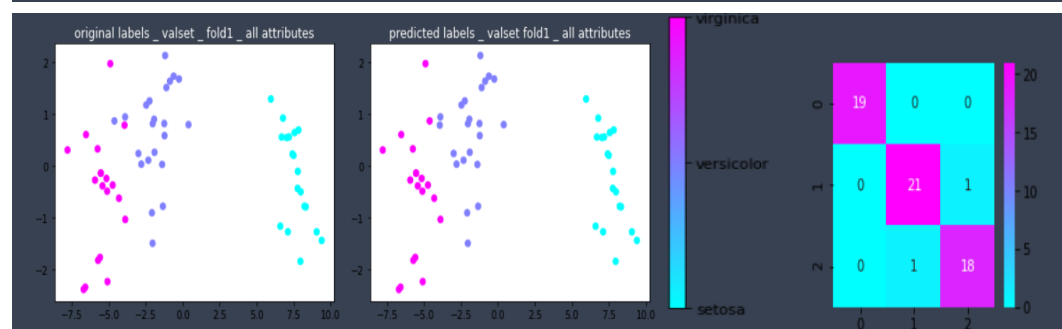
Data=validation set

test_size=0.2

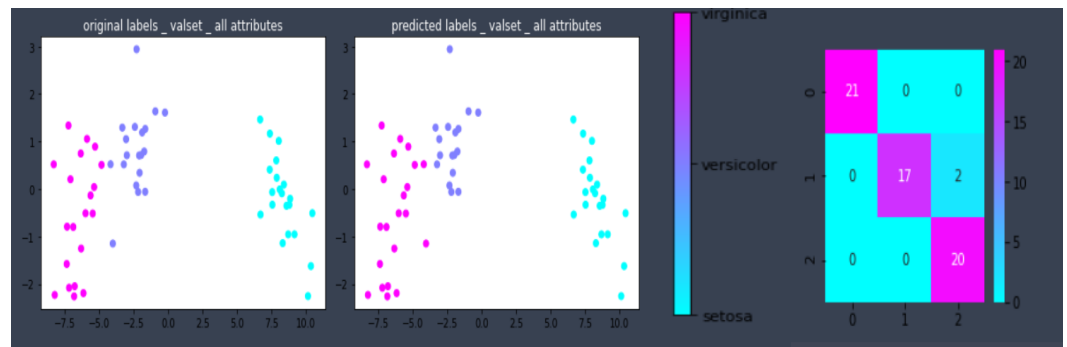
fold=1

accuracy =96%

Data=validation set

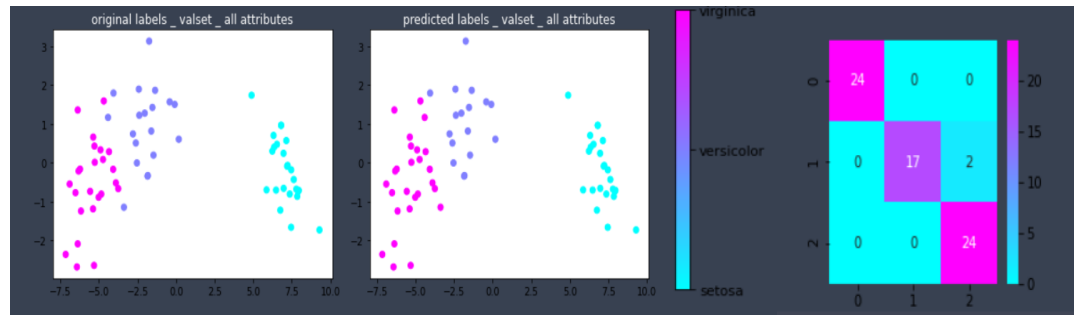


test_size=0.2
fold=2
accuracy =96%



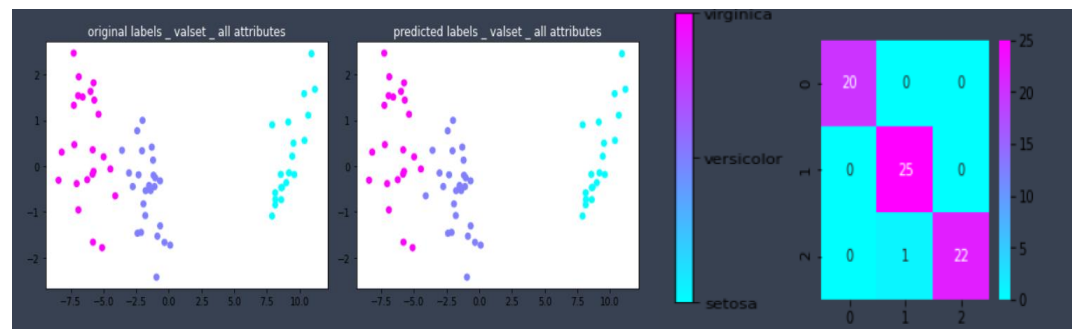
Data=validation set

test_size=0.1
fold=1
accuracy =96%



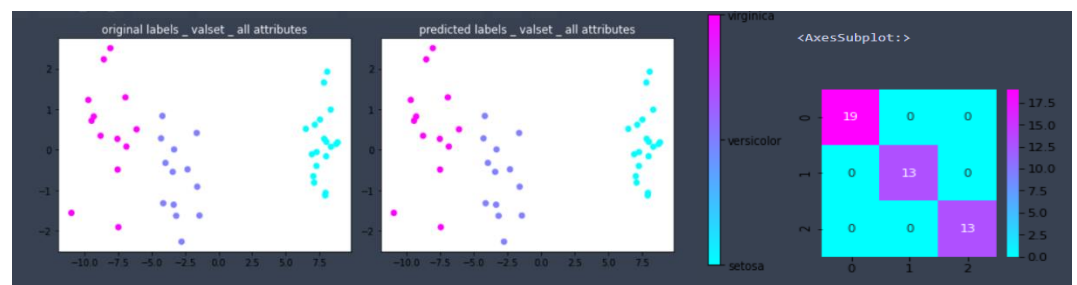
Data=validation set

test_size=0.1
fold=2
accuracy =96%



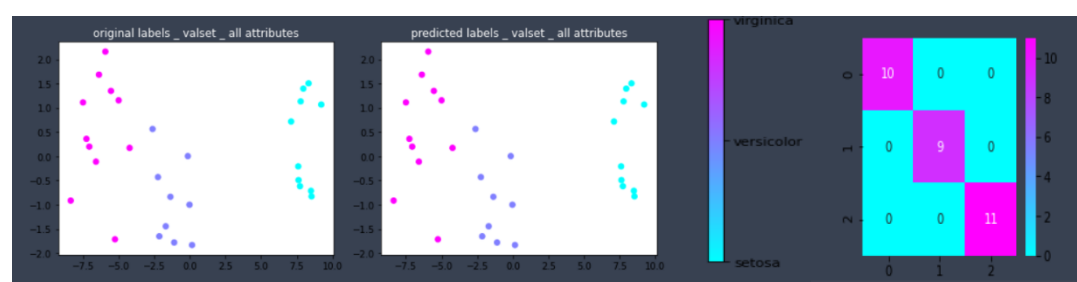
Data=test set

test_size=0.3
accuracy =100%



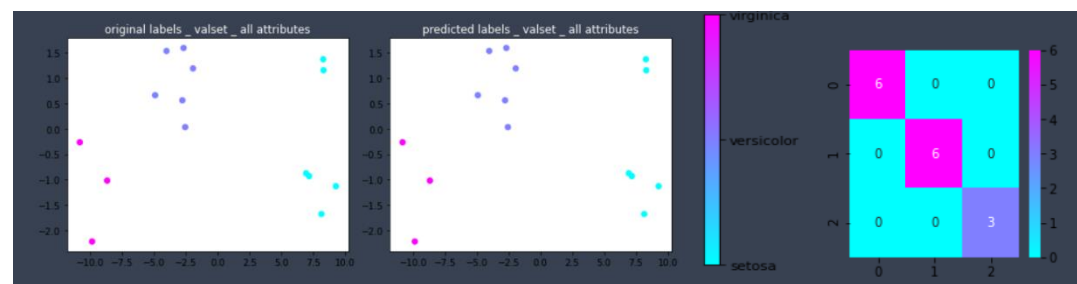
Data=test set

test_size=0.2
accuracy =100%



Data=test set

test_size=0.1
accuracy =100%

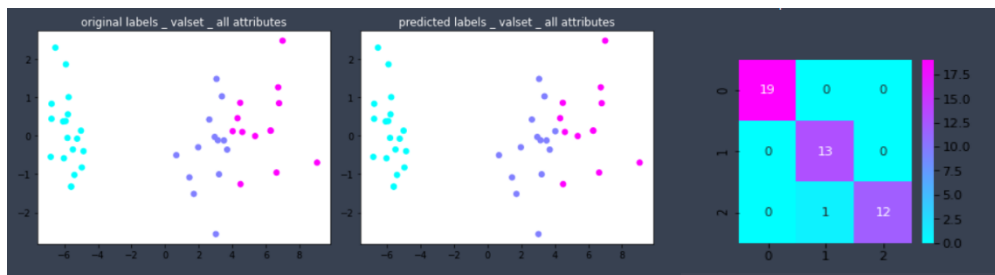


داده ها با ویژگی های کمتر:

Att={1,2,3}

Acc=96%

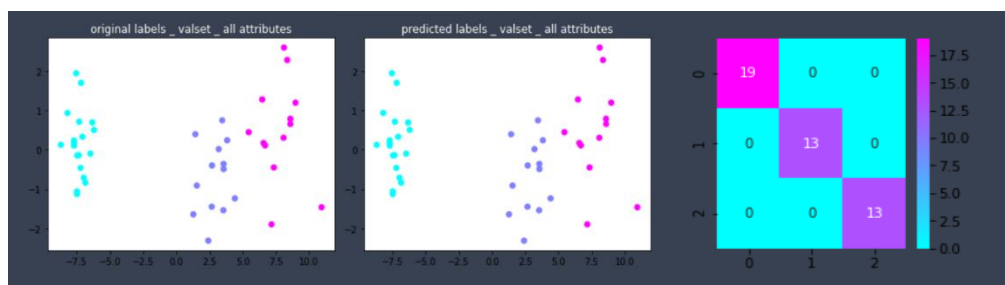
Data=test set



Att={2,3,4}

Acc=100%

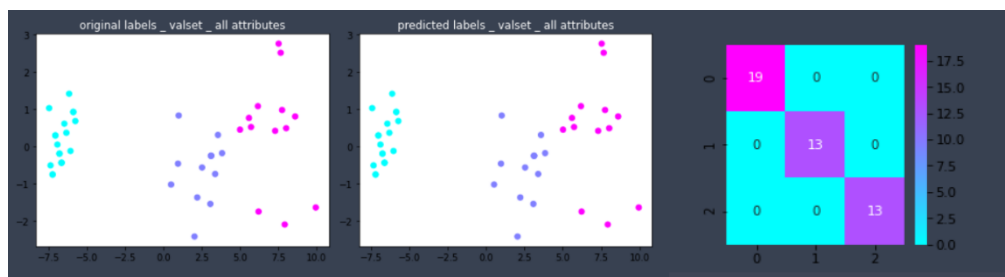
Data=test set



Att={2,3}

Acc=100%

Data=test set



نتیجه این قسمت : lda به خوبی می تواند داده های Iris را دسته بندی کند و بهترین ویژگی های جدا کننده این دیتاست ویژگی های 3 و 4 هستند.

ب:

در قسمت دوم با داشتن مقدار سمپل ها، پارامترهای گوسی را به دست می آوریم تا با الگوریتم بیز دسته بندی را انجام دهیم.

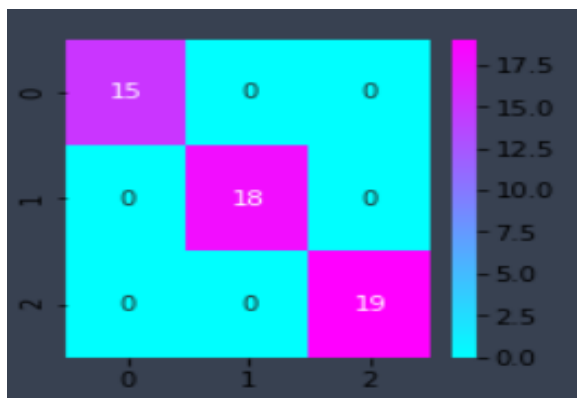
ماتریس کواریانس کلاس ها:

میانگین کلاس ها:

```
cov class 0
[[0.0725    0.0915625  0.0075    0.0015625 ]
 [0.0915625  0.17058594 0.01734375 0.00269531]
 [0.0075     0.01734375 0.025625   0.00390625]
 [0.0015625  0.00269531 0.00390625 0.01589844]]
cov class 1
[[0.2421    0.06815   0.20685   0.066    ]
 [0.06815   0.056475   0.101275  0.046    ]
 [0.20685   0.101275   0.323475  0.1155   ]
 [0.066     0.046     0.1155    0.057    ]]
cov class 2
[[0.33397924 0.12705882 0.30359862 0.0617301 ]
 [0.12705882 0.15411765 0.13882353 0.08823529]
 [0.30359862 0.13882353 0.34525952 0.09148789]
 [0.0617301  0.08823529 0.09148789 0.08013841]]
```

	0	1	2	3
class				
0.0	5.06	3.180000	1.500000	0.300000
1.0	5.04	3.500000	1.560000	0.280000
2.0	5.05	3.516667	1.516667	0.233333

از روی میانگین و ماتریس کواریانس تابع دسته بندیز را برای هر کلاس به دست آوردیم. و بر روی داده های تست اعمال کردیم.



Confusion ماتریس برای برچسب دیتاهای پیشبینی شده به صورت

روبرو است. و دقت دسته بند بر روی داده های تست 100% است.

این کار را با تغییر fold و تغییر اندازه test تکرار کردیم و نتیجه تغییری نکرد.

Accuracy=100%

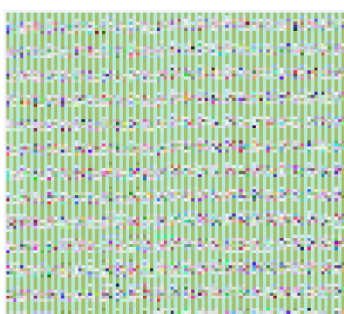
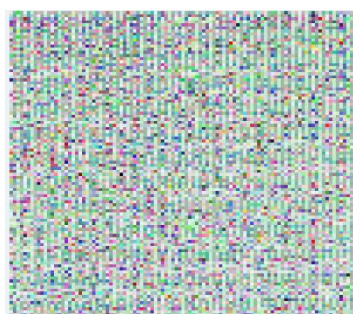
قسمت دوم:

در این قسمت بعد از بارگذاری تصاویر میوه ها ، آنها را به ماتریس تبدیل کردیم و سپس به بردار یک بعدی تبدیل کرده و تمام میوه ها را به عنوان ستون های دیتافریم قرار دادیم. با به دست آوردن میانگین داده ها و مقدار ویژه ها و بردار ویژه ها بر روی داده ها pca اجرا کردیم و بعد از کاهش بعد چند نمونه از دیتا ها را دوباره بازسازی کردیم. همانطور که مشخص است pca بر روی داده های متفاوت نمیتواند تصویر دقیقی از آنها بازسازی کند. چون تصاویر متفاوت بوده اند. اما ویژگی هایی در تصاویر پیدا کرده که با آنها می تواند دسته بندی را انجام دهد. ولی به لحاظ دیداری چندان قابل تشخیص برای انسان نیست. و تغییر تعداد بردار ویژه ها تفاوت چشم گیری ایجاد نکرد.

کامکوات

ذرت

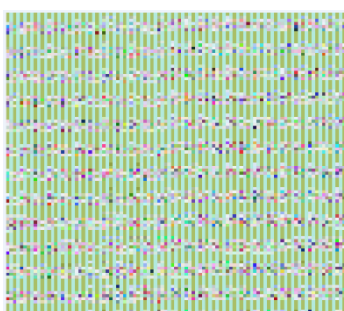
موز



باز سازی

تصویر با 10

بردار ویژه اول



باز سازی

تصویر با

1000

بردار ویژه اول