# WEEK-7

## Linear Discriminanat Analysis

Linear Discriminant Analysis (LDA) is a supervised learning algorithm used for classification tasks in machine learning. It is a technique used to find a linear combination of features that best separates the classes in a dataset. LDA works by projecting the data onto a lower-dimensional space that maximizes the separation between the classes. It does this by finding a set of linear discriminants that maximize the ratio of between-class variance to within-class variance. In other words, it finds the directions in the feature space that best separate the different classes of data.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sn
from sklearn.datasets import load_wine

wine_info=load_wine()
X=pd.DataFrame(wine_info.data,columns=wine_info.feature_names)

X.head()
```

| | alcohol | malic_acid | ash | alcalinity_of_ash | magnesium | total_phenols | \ |
|---|---|---|---|---|---|---|---|
| 0 | 14.23 | 1.71 | 2.43 | 15.6 | 127.0 | 2.80 | |
| 1 | 13.20 | 1.78 | 2.14 | 11.2 | 100.0 | 2.65 | |
| 2 | 13.16 | 2.36 | 2.67 | 18.6 | 101.0 | 2.80 | |
| 3 | 14.37 | 1.95 | 2.50 | 16.8 | 113.0 | 3.85 | |
| 4 | 13.24 | 2.59 | 2.87 | 21.0 | 118.0 | 2.80 | |

| | flavanoids | nonflavanoid_phenols | proanthocyanins | color_intensity | hue | \ |
|---|---|---|---|---|---|---|
| 0 | 3.06 | 0.28 | 2.29 | 5.64 | 1.04 | |
| 1 | 2.76 | 0.26 | 1.28 | 4.38 | 1.05 | |
| 2 | 3.24 | 0.30 | 2.81 | 5.68 | 1.03 | |
| 3 | 3.49 | 0.24 | 2.18 | 7.80 | 0.86 | |
| 4 | 2.69 | 0.39 | 1.82 | 4.32 | 1.04 | |

| | od280/od315_of_diluted_wines | proline |
|---|---|---|
| 0 | 3.92 | 1065.0 |
| 1 | 3.40 | 1050.0 |
| 2 | 3.17 | 1185.0 |
| 3 | 3.45 | 1480.0 |
| 4 | 2.93 | 735.0 |

```
X.shape
```

(178, 13)

```
X.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177

```
Data columns (total 13 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   alcohol                       178 non-null    float64
 1   malic_acid                    178 non-null    float64
 2   ash                           178 non-null    float64
 3   alcalinity_of_ash             178 non-null    float64
 4   magnesium                     178 non-null    float64
 5   total_phenols                 178 non-null    float64
 6   flavanoids                    178 non-null    float64
 7   nonflavanoid_phenols          178 non-null    float64
 8   proanthocyanins               178 non-null    float64
 9   color_intensity               178 non-null    float64
 10  hue                           178 non-null    float64
 11  od280/od315_of_diluted_wines  178 non-null    float64
 12  proline                       178 non-null    float64
dtypes: float64(13)
memory usage: 18.2 KB
```

X.describe()

| | alcohol | malic_acid | ash | alcalinity_of_ash | magnesium |
|---|---|---|---|---|---|
| count | 178.000000 | 178.000000 | 178.000000 | 178.000000 | 178.000000 |
| mean | 13.000618 | 2.336348 | 2.366517 | 19.494944 | 99.741573 |
| std | 0.811827 | 1.117146 | 0.274344 | 3.339564 | 14.282484 |
| min | 11.030000 | 0.740000 | 1.360000 | 10.600000 | 70.000000 |
| 25% | 12.362500 | 1.602500 | 2.210000 | 17.200000 | 88.000000 |
| 50% | 13.050000 | 1.865000 | 2.360000 | 19.500000 | 98.000000 |
| 75% | 13.677500 | 3.082500 | 2.557500 | 21.500000 | 107.000000 |
| max | 14.830000 | 5.800000 | 3.230000 | 30.000000 | 162.000000 |

| | total_phenols | flavanoids | nonflavanoid_phenols | proanthocyanins |
|---|---|---|---|---|
| count | 178.000000 | 178.000000 | 178.000000 | 178.000000 |
| mean | 2.295112 | 2.029270 | 0.361854 | 1.590899 |
| std | 0.625851 | 0.998859 | 0.124453 | 0.572359 |
| min | 0.980000 | 0.340000 | 0.130000 | 0.410000 |
| 25% | 1.742500 | 1.205000 | 0.270000 | 1.250000 |
| 50% | 2.355000 | 2.135000 | 0.340000 | 1.555000 |
| 75% | 2.800000 | 2.875000 | 0.437500 | 1.950000 |
| max | 3.880000 | 5.080000 | 0.660000 | 3.580000 |

| | color_intensity | hue | od280/od315_of_diluted_wines | proline |
|---|---|---|---|---|
| count | 178.000000 | 178.000000 | 178.000000 | 178.000000 |
| mean | 5.058090 | 0.957449 | 2.611685 | 746.893258 |
| std | 2.318286 | 0.228572 | 0.709990 | 314.907474 |
| min | 1.280000 | 0.480000 | 1.270000 | 278.000000 |
| 25% | 3.220000 | 0.782500 | 1.937500 | 500.500000 |
| 50% | 4.690000 | 0.965000 | 2.780000 | 673.500000 |
| 75% | 6.200000 | 1.120000 | 3.170000 | 985.000000 |
| max | 13.000000 | 1.710000 | 4.000000 | 1680.000000 |

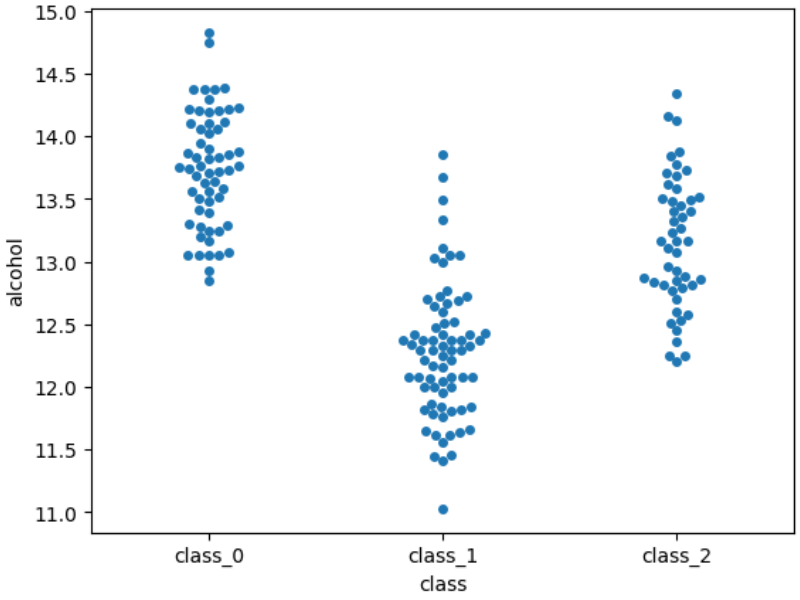
```
Y=pd.Categorical.from_codes(wine_info.target,wine_info.target_names)

Y.unique()

['class_0', 'class_1', 'class_2']
Categories (3, object): ['class_0', 'class_1', 'class_2']

df=X.join(pd.Series(Y,name="class"))

df.head()
```

| | alcohol | malic_acid | ash | alcalinity_of_ash | magnesium | total_phenols | \ |
|---|---|---|---|---|---|---|---|
| 0 | 14.23 | 1.71 | 2.43 | 15.6 | 127.0 | 2.80 | |
| 1 | 13.20 | 1.78 | 2.14 | 11.2 | 100.0 | 2.65 | |
| 2 | 13.16 | 2.36 | 2.67 | 18.6 | 101.0 | 2.80 | |
| 3 | 14.37 | 1.95 | 2.50 | 16.8 | 113.0 | 3.85 | |
| 4 | 13.24 | 2.59 | 2.87 | 21.0 | 118.0 | 2.80 | |

| | flavanoids | nonflavanoid_phenols | proanthocyanins | color_intensity | hue | \ |
|---|---|---|---|---|---|---|
| 0 | 3.06 | 0.28 | 2.29 | 5.64 | 1.04 | |
| 1 | 2.76 | 0.26 | 1.28 | 4.38 | 1.05 | |
| 2 | 3.24 | 0.30 | 2.81 | 5.68 | 1.03 | |
| 3 | 3.49 | 0.24 | 2.18 | 7.80 | 0.86 | |
| 4 | 2.69 | 0.39 | 1.82 | 4.32 | 1.04 | |

| | od280/od315_of_diluted_wines | proline | class |
|---|---|---|---|
| 0 | 3.92 | 1065.0 | class_0 |
| 1 | 3.40 | 1050.0 | class_0 |
| 2 | 3.17 | 1185.0 | class_0 |
| 3 | 3.45 | 1480.0 | class_0 |
| 4 | 2.93 | 735.0 | class_0 |

```
sn.swarmplot(x=df["class"],y=df["alcohol"])
```



```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y_labelled=le.fit_transform(Y)
y_labelled[0:5]
```

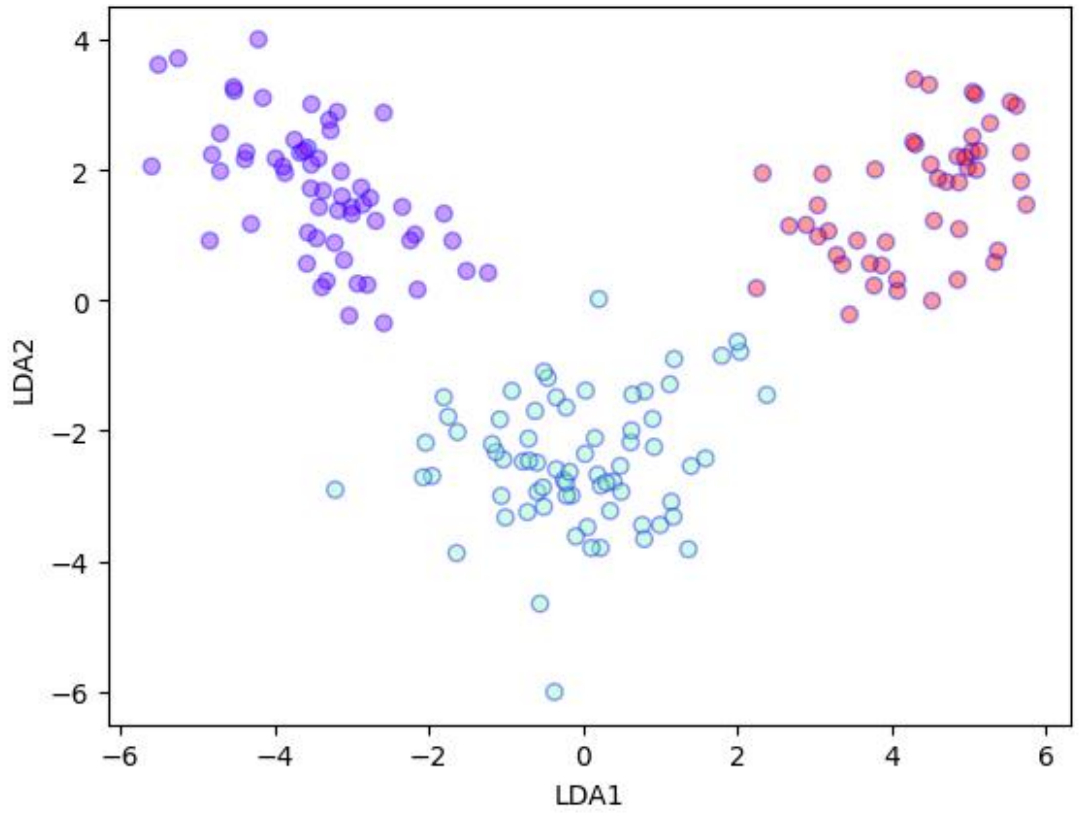
```
array([0, 0, 0, 0, 0])
```

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.model_selection import train_test_split

lda=LinearDiscriminantAnalysis()
X_lda=lda.fit_transform(X,y_labelled)
lda.explained_variance_ratio_
```

```
array([0.68747889, 0.31252111])
```

```
plt.xlabel("LDA1")
plt.ylabel("LDA2")
plt.scatter(
    X_lda[:,0],
    X_lda[:,1],
    c=y_labelled,
    cmap="rainbow",
    alpha=0.4,
    edgecolors='b'
)
```

```
<matplotlib.collections.PathCollection at 0x1b1c890a500>
```




```
X_train,X_test,Y_train,Y_test=train_test_split(X_lda,y_labelled,random_state=20,test_size=0.4)

from sklearn.tree  import DecisionTreeClassifier

model=DecisionTreeClassifier()
model.fit(X_train,Y_train)
model.score(X_test,Y_test)
```

0.9861111111111112