Linear Discriminant Analysis

February 23, 2022

1 Linear Discremenrt Analyss

```python
[1]: import numpy as np
     import pandas as pd
     from sklearn import datasets
```

```python
[2]: path_in = "/Users/saeedkhalili/Desktop/IntroML_Grant/iris.csv"
```

```python
[3]: df = pd.read_csv(path_in)
```

```python
[4]: df.head()
```

```
[4]:    sepal_length  sepal_width  petal_length  petal_width species
     0           5.1          3.5           1.4          0.2  setosa
     1           4.9          3.0           1.4          0.2  setosa
     2           4.7          3.2           1.3          0.2  setosa
     3           4.6          3.1           1.5          0.2  setosa
     4           5.0          3.6           1.4          0.2  setosa
```

```python
[5]: df.tail()
```

```
[5]:      sepal_length  sepal_width  petal_length  petal_width    species
     145           6.7          3.0           5.2          2.3  virginica
     146           6.3          2.5           5.0          1.9  virginica
     147           6.5          3.0           5.2          2.0  virginica
     148           6.2          3.4           5.4          2.3  virginica
     149           5.9          3.0           5.1          1.8  virginica
```

```python
[6]: df.shape
```

```
[6]: (150, 5)
```

```python
[7]: df.size
```

```
[7]: 750
```

```python
[8]: df.describe()
```

```
[8]:         sepal_length   sepal_width   petal_length   petal_width
      count     150.000000    150.000000     150.000000    150.000000
      mean        5.843333      3.057333       3.758000      1.199333
      std         0.828066      0.435866       1.765298      0.762238
      min         4.300000      2.000000       1.000000      0.100000
      25%         5.100000      2.800000       1.600000      0.300000
      50%         5.800000      3.000000       4.350000      1.300000
      75%         6.400000      3.300000       5.100000      1.800000
      max         7.900000      4.400000       6.900000      2.500000
```

```
[9]: y = df['species']
```

```
[10]: y
```

```
[10]: 0        setosa
      1        setosa
      2        setosa
      3        setosa
      4        setosa
                ...
      145    virginica
      146    virginica
      147    virginica
      148    virginica
      149    virginica
      Name: species, Length: 150, dtype: object
```

```
[11]: from sklearn.preprocessing import LabelEncoder
      y = LabelEncoder().fit_transform(y)
```

```
[12]: y
```

```
[12]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
             0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
             0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
             1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
             1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
             2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
             2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

```
[13]: X = df.drop(['species'], axis = 1)
```

```
[14]: from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
      lda = LinearDiscriminantAnalysis()
```

```
[15]: lda.fit(X, y)
```

```
[15]: LinearDiscriminantAnalysis()
```

```
[16]: yhat = lda.predict(X)
```

```
[17]: yhat
```

```
[17]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
              0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
              0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
              1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,
              1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
              2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
              2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

```
[18]: from sklearn import metrics
```

```
[19]: metrics.accuracy_score(y, yhat)
```

```
[19]: 0.98
```

```
[20]: from sklearn.metrics import confusion_matrix
      confusion_matrix(y, yhat)
```

```
[20]: array([[50,  0,  0],
             [ 0, 48,  2],
             [ 0,  1, 49]])
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```