Chapter 7

Gaussian Discriminant Analysis

Discriminant Functions

- Consider a classifier built using functions

Discriminant Functions

- Consider a classifier built using functions

$$g_i(\bar{\mathbf{x}}), i = 1,2, \ldots, L$$

Discriminant Functions

- The input vector $\bar{\mathbf{x}}$ is assigned to class l if

Discriminant Functions

- The **input vector** $\bar{\mathbf{x}}$ is assigned to class l if

$$g_l(\bar{\mathbf{x}}) = \arg\max_{1 \le i \le L} g_i(\bar{\mathbf{x}})$$

Discriminant Functions

- These $g_i(\bar{\mathbf{x}})$ are termed

  ____

Discriminant Functions

- These $g_i(\bar{\mathbf{x}})$ are termed <u>discriminant functions</u>

Gaussian Density

- Recall, the expression for the *Gaussian PDF* is

Gaussian Density

- Recall, the expression for the *Gaussian PDF* is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Gaussian Density

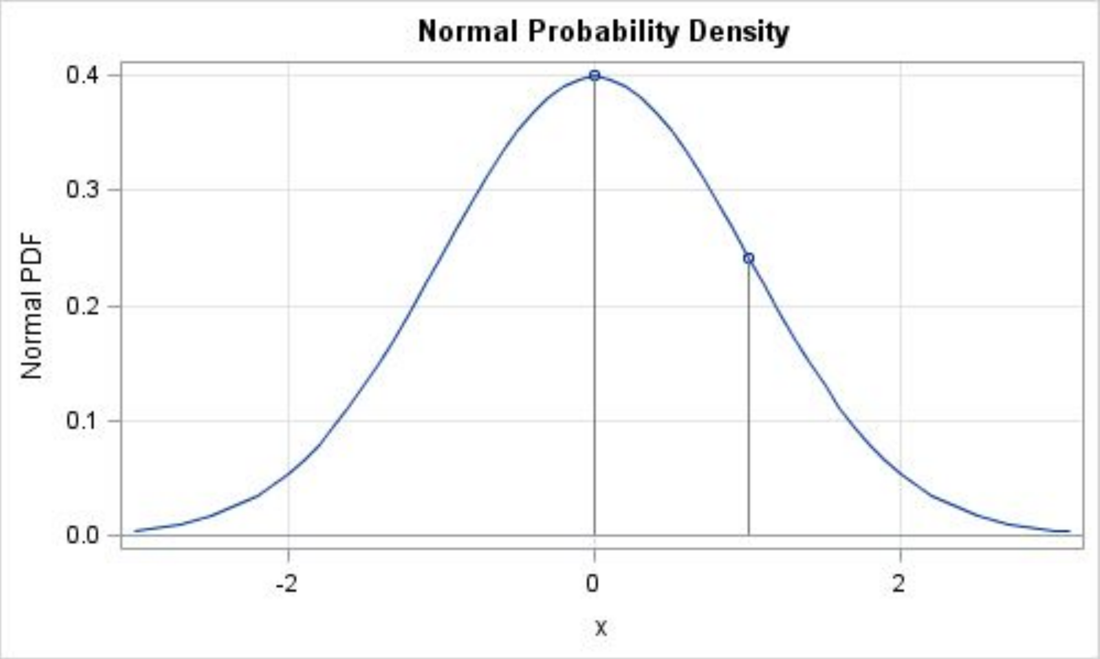
- The *mean* and *variance* of the Gaussian RV are

Gaussian Density

- The *mean* and *variance* of the Gaussian RV are

$$E\{X\} = \mu$$
$$E\{(X - \mu)^2\} = \sigma^2$$

Gaussian Density



Multivariate Gaussian Density

- Recall, the PDF of a *Gaussian random vector* is given as

Multivariate Gaussian Density

- Recall, the PDF of a *Gaussian random vector* is given as

$$f_{\bar{\mathbf{x}}}(\bar{\mathbf{x}}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{R}|}} e^{-\frac{1}{2}(\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}})^T\mathbf{R}^{-1}(\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}})}$$

Multivariate Gaussian Density

- The *mean* and *covariance matrix* are defined as

Multivariate Gaussian Density

- The *mean* and *covariance matrix* are defined as

$$E\{\bar{\mathbf{x}}\} = \bar{\boldsymbol{\mu}}$$

$$E\{(\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}})^T\} = \mathbf{R}$$

Gaussian Discriminant Analysis

- Consider the input vectors $\bar{\mathbf{x}}$ drawn from *two Gaussian classes*

$\mathcal{C}_0$:

$\mathcal{C}_1$:

- Prior probabilities p_0, p_1

Gaussian Discriminant Analysis

- Consider the input vectors $\bar{\mathbf{x}}$ drawn from *two Gaussian classes*

$$\mathcal{C}_0: \text{Mean } \bar{\boldsymbol{\mu}}_0 \text{ and covariance } \mathbf{R}$$
$$\mathcal{C}_1: \text{Mean } \bar{\boldsymbol{\mu}}_1 \text{ and covariance } \mathbf{R}$$

- Prior probabilities p_0, p_1

Gaussian Discriminant Analysis

- Also termed Gaussian Discriminant Analysis

Gaussian Discriminant Analysis

- Thus, the *likelihoods* of the two classes are

$$p(\bar{\mathbf{x}}; \mathcal{C}_0) =$$

$$p(\bar{\mathbf{x}}; \mathcal{C}_1) =$$

Gaussian Discriminant Analysis

- Thus, the *likelihoods* of the two classes are

$$p(\bar{\mathbf{x}}; \mathcal{C}_0) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{R}|}} e^{-\frac{1}{2}(\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} (\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_0)}$$

$$p(\bar{\mathbf{x}}; \mathcal{C}_1) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{R}|}} e^{-\frac{1}{2}(\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_1)^T \mathbf{R}^{-1} (\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_1)}$$

Maximum Likelihood rule

- Choose the class that *maximizes the posterior probability*

Maximum Likelihood rule

- Therefore, **choose** $\mathcal{C}_0$ if

$$p_0 \times p(\bar{\mathbf{x}}; \mathcal{C}_0) \geq p_1 \times p(\bar{\mathbf{x}}; \mathcal{C}_1)$$

Maximum Likelihood rule

- Therefore, choose $\mathcal{C}_0$ if

$$p_0 \times p(\bar{\mathbf{x}}; \mathcal{C}_0) \geq p_1 \times p(\bar{\mathbf{x}}; \mathcal{C}_1)$$

$$\Rightarrow \frac{p_0}{\sqrt{(2\pi)^n |\mathbf{R}|}} e^{-\frac{1}{2}(\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} (\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_0)} \geq \frac{p_1}{\sqrt{(2\pi)^n |\mathbf{R}|}} e^{-\frac{1}{2}(\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_1)^T \mathbf{R}^{-1} (\bar{\mathbf{x}}-\bar{\boldsymbol{\mu}}_1)}$$

$$\Rightarrow \ln p_0 - \frac{1}{2} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_0)$$

$$\geq \ln \mathrm{p}_1 - \frac{1}{2} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_1)^T \mathbf{R}^{-1} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_1)$$

Maximum Likelihood rule

- This ***discriminant function*** can be simplified as

Choose $\mathcal{C}_0$:

Choose $\mathcal{C}_1$:

Maximum Likelihood rule

- This *discriminant function* can be simplified as

Choose $\mathcal{C}_0$: $\bar{\mathbf{h}}^T(\bar{\mathbf{x}} - \tilde{\boldsymbol{\mu}}) \geq \ln \frac{p_1}{p_0}$

Choose $\mathcal{C}_1$: $\bar{\mathbf{h}}^T(\bar{\mathbf{x}} - \tilde{\boldsymbol{\mu}}) < \ln \frac{p_1}{p_0}$

Maximum Likelihood rule

- where

$$\tilde{\boldsymbol{\mu}} =$$

$$\bar{\mathbf{h}} =$$

Maximum Likelihood rule

- where

$$\tilde{\boldsymbol{\mu}} = \frac{1}{2}(\bar{\boldsymbol{\mu}}_0 + \bar{\boldsymbol{\mu}}_1)$$

$$\bar{\mathbf{h}} = \mathbf{R}^{-1}(\bar{\boldsymbol{\mu}}_0 - \bar{\boldsymbol{\mu}}_1)$$

Linear classifier

- Thus, the classifier is *linear*
- It is characterized by the *hyperplane*

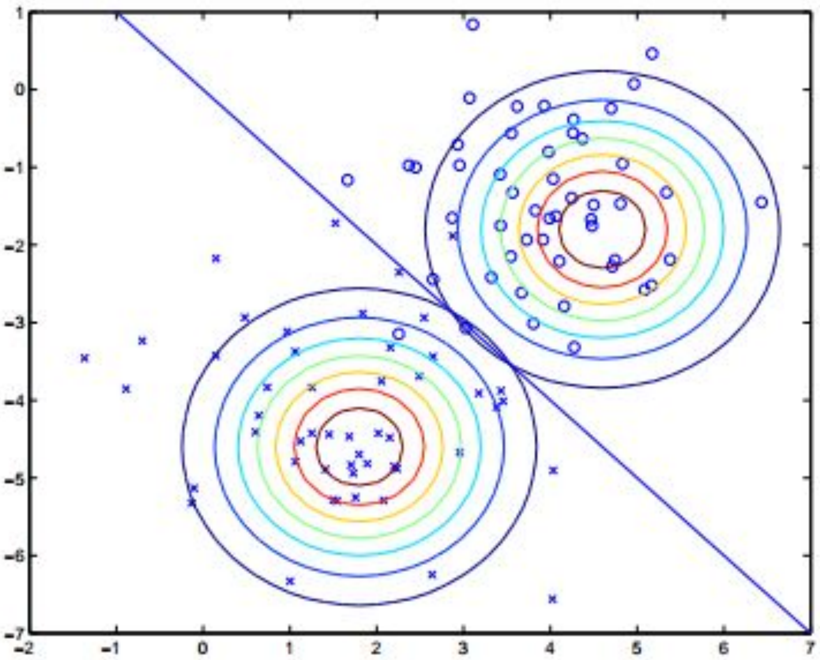
Linear classifier

- Thus, the classifier is ***linear***
- It is characterized by the ***hyperplane***

$$\bar{\mathbf{h}}^T(\bar{\mathbf{x}} - \widetilde{\boldsymbol{\mu}}) = \ln\frac{p_1}{p_0}$$

- Proof in Appendix

Gaussian Discriminant Classifier



```python
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
```

```python
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
```

```python
DiabetesData = pd.read_csv('Diabetes.csv')
```

glucose	bloodpres	diabetes
40	85	0
40	92	0
45	63	1
45	80	0
40	73	1
45	82	0
40	85	0
30	63	1
65	65	1
45	82	0
35	73	1
45	90	0
50	68	1
40	93	0
35	80	1
50	70	1
40	73	1
40	67	1
40	75	1
40	80	1
40	72	1
40	88	0
40	78	1
45	98	0
40	88	0
60	67	1
40	85	0
40	88	0
45	78	0

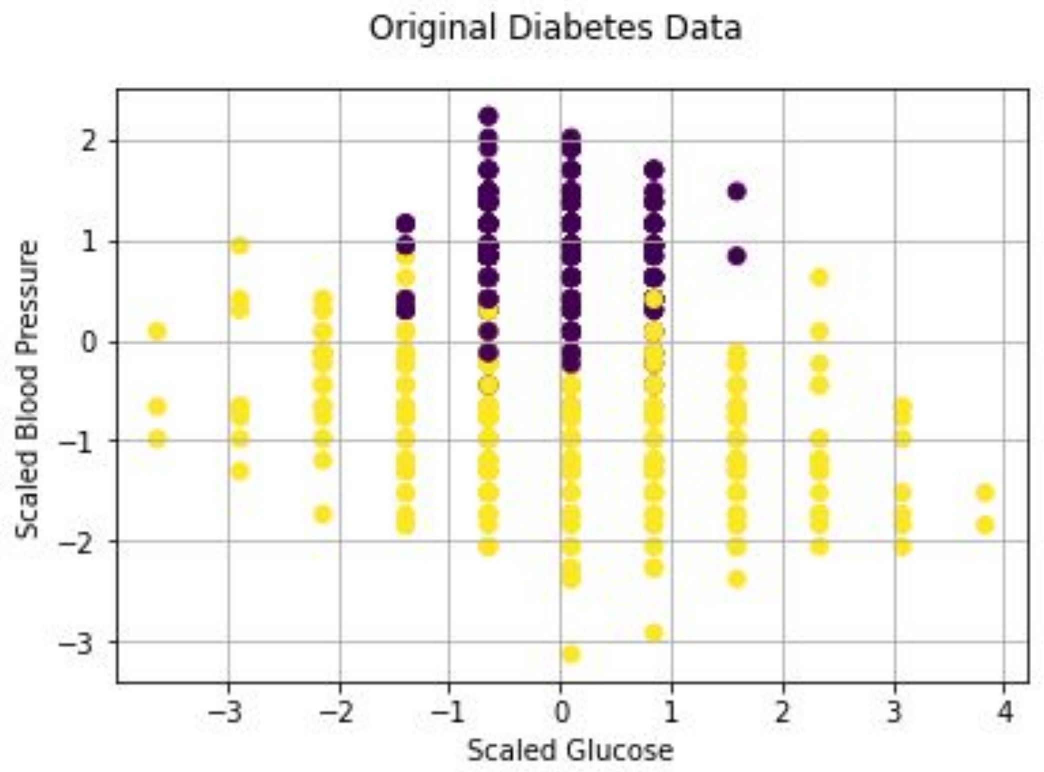
Diabetes

```
X = DiabetesData.iloc[:, [0, 1]].values
Y = DiabetesData.iloc[:, 2].values
```

```
scaler = StandardScaler();
X = scaler.fit_transform(X)
```

```
Xtrain, Xtest, Ytrain, Ytest \
= train_test_split(X, Y, test_size = 0.20, random_state = 5)
```

```python
plt.figure(1);
plt.scatter(X[:, 0], X[:, 1], c = Y)
plt.suptitle('Original Diabetes Data')
plt.xlabel('Scaled Glucose')
plt.ylabel('Scaled Blood Pressure')
plt.grid(1,which='both')
plt.axis('tight')
plt.show()
```


Original Diabetes Data
Scaled Blood Pressure
Scaled Glucose
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
4

```
lda = LinearDiscriminantAnalysis()
lda.fit(Xtrain,Ytrain)
Y_pred = lda.predict(X)
```

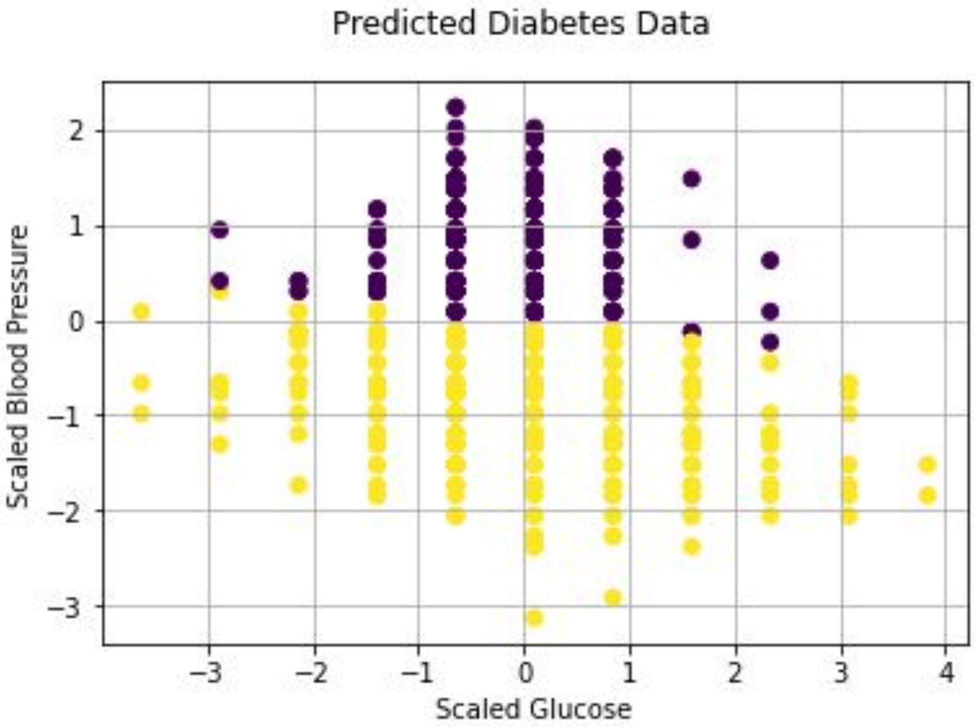
```python

ldascore = accuracy_score(lda.predict(Xtest),Ytest)
print('Accuracy score of LDA Classifier is',100*ldascore,'%\n')

```

```
Accuracy score of LDA Classifier is 94.47236180904522 %
```

```python
plt.figure(1);
plt.scatter(X[:, 0], X[:, 1], c = Y_pred)
plt.suptitle('Predicted Diabetes Data')
plt.xlabel('Scaled Glucose')
plt.ylabel('Scaled Blood Pressure')
plt.grid(1,which='both')
plt.axis('tight')
plt.show()
```



Predicted Diabetes Data
Scaled Blood Pressure
Scaled Glucose
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
4

Instructors may use this white area (14.5 cm / 25.4 cm) for the text. Three options provided below for the font size.

Font: Avenir (Book), Size: 32, Colour: Dark Grey

Font: Avenir (Book), Size: 28, Colour: Dark Grey

Font: Avenir (Book), Size: 24, Colour: Dark Grey

Do not use the space below.

Appendix: Proof of LDA

Proof of LDA

- The classifier chooses $\boldsymbol{C_0}$ if

$$(\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_0) - (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_1)^T \mathbf{R}^{-1} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}_1) \le 2 \ln \frac{p_0}{p_1}$$

$$\Rightarrow 2(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} \bar{\mathbf{x}} + \bar{\boldsymbol{\mu}}_0^T \mathbf{R}^{-1} \bar{\boldsymbol{\mu}}_0 - \bar{\boldsymbol{\mu}}_1 \mathbf{R}^{-1} \bar{\boldsymbol{\mu}}_1 \le 2 \ln \frac{p_0}{p_1}$$

Proof of LDA

- The classifier chooses $\boldsymbol{C_0}$ if

$$\Rightarrow (\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} \bar{\mathbf{x}} - \frac{1}{2} (\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} (\bar{\boldsymbol{\mu}}_1 + \bar{\boldsymbol{\mu}}_0) \leq \ln \frac{p_0}{p_1}$$

$$\Rightarrow (\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_0)^T \mathbf{R}^{-1} \left(\bar{\mathbf{x}} - \frac{1}{2} (\bar{\boldsymbol{\mu}}_1 + \bar{\boldsymbol{\mu}}_0) \right) \leq \ln \frac{p_0}{p_1}$$

$$\Rightarrow (\bar{\boldsymbol{\mu}}_0 - \bar{\boldsymbol{\mu}}_1)^T \mathbf{R}^{-1} \left(\bar{\mathbf{x}} - \frac{1}{2} (\bar{\boldsymbol{\mu}}_1 + \bar{\boldsymbol{\mu}}_0) \right) \geq \ln \frac{p_1}{p_0}$$