Ass - 1

(1) Mean, $\mu = 42$ months

Std. Dev, $\sigma = 8$ months

For normal dis., $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$$f_X(x) = \frac{1}{8\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-42)^2}{64}}$$

$$P(20 < X < 30) = \int_{20}^{30} f_X(x)\,dx$$

$$= \frac{1}{8\sqrt{2\pi}} \int_{20}^{30} e^{-\frac{(x-42)^2}{128}}\,dx$$

$$= 0.06338$$

(2) A meaningful measure of location that can be calculated for this type of data is median.

Sorted order: 36, 45, 51, 63, 75, 80, 90, 100+

$n = 8 \Rightarrow r = \frac{n}{2} = 4$

$$\text{Median} = \frac{x_4 + x_5}{2} = \frac{63 + 75}{2} = 69$$

(3) Case 1 :- 1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91 $(N = 8)$

$$\text{Mean}, \bar{x} = \frac{1}{8}\sum_{n=1}^{8} x_n = \frac{19.32}{8} = \boxed{2.415 \text{ sec}}$$

$$\text{Variance}, s_x^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \bar{x})^2 = \frac{1}{7}\sum_{n=1}^{8}(x_n - 2.415)^2 = \frac{1.9976}{7}$$

Std. dev, $s_x = \sqrt{0.285} = \boxed{0.534}$ $= \boxed{0.285}$

Case 2 :- 1.83, 1.99, 3.13, 3.29, 2.65, 2.87, 3.40, 2.46, 1.89, 3

$N = 10$

Mean, $\bar{x} = \frac{1}{10} \sum_{n=1}^{10} x_n = \frac{27.86}{10} =$ **2.786 sec**

Variance, $s_x^2 = \frac{1}{9} \sum_{n=1}^{10} (x_n - \bar{x})^2 = \frac{3.548}{9} =$ **0.3942**

Std. dev., $s_x = \sqrt{0.3942} =$ **0.628**

(4) (a) Normalized weight = $\frac{\text{Weight} - \text{Min. weight}}{\text{Max. weight} - \text{Min. weight}}$

Min. weight = 41

Max. weight = 136

(3) Interpretation :-

→ Gasoline B has a higher median cold start time

→ Gasoline B shows a wider interquartile range, indicating more variability.

→ Gasoline B shows higher & more variable ignition times, which may affect performance during cold starts

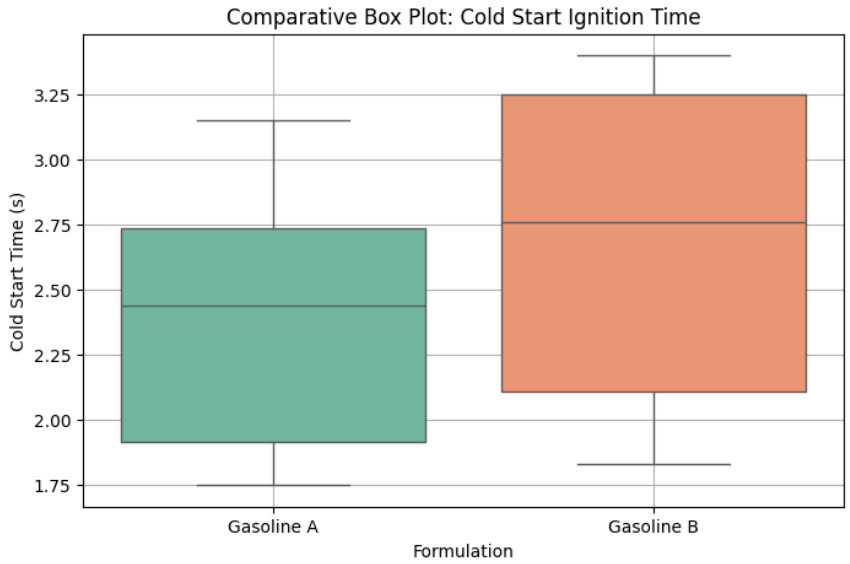
→ Therefore, Gasoline A tends to start the vehicle faster & more consistently.

Question 1: Output

```
Probability (by integration) that X is between 20 and 30: 0.0638
```

Question 3: Output

```
Case 1:
Mean = 2.42, Variance = 0.2854, Std Dev = 0.5342

Case 2:
Mean = 2.69, Variance = 0.3833, Std Dev = 0.6191
```

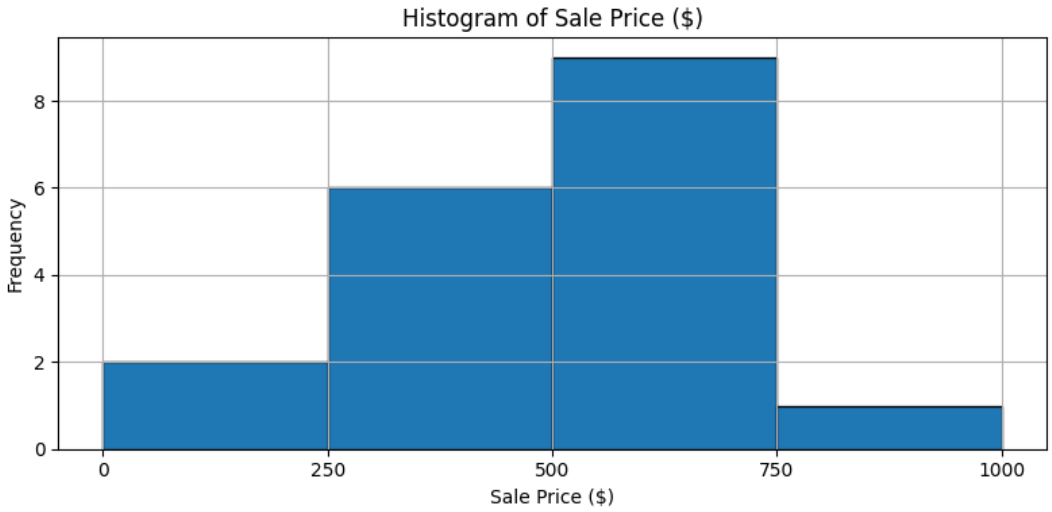


Question 4: Output

	Name	Weight (kg)	Height (m)	Normalized Weight	Weight Category	BMI
0	P. Lee	50	1.52	0.094737	Low	21.641274
1	R. Jones	115	1.77	0.778947	High	36.707204
2	J. Smith	96	1.83	0.578947	Medium	28.666129
3	A. Patel	41	1.55	0.000000	Low	17.065557
4	M. Owen	79	1.82	0.400000	Medium	23.849777
5	S. Green	109	1.89	0.715789	High	30.514263
6	N. Cook	73	1.76	0.336842	Medium	23.566632
7	W. Hands	104	1.71	0.663158	High	35.566499
8	P. Rice	64	1.74	0.242105	Medium	21.138856
9	F. Marsh	136	1.78	1.000000	High	42.923873

Question 5: Output

(a)



(b)

Contingency Table:

Store	Product Category	Desktop	Laptop	Printer	Scanner
New York, NY		3	1	2	4
Washington, DC		2	2	2	2

(c)

Summary by Customer:

Customer	Count	Total_Sales
B. March	3	1700
E. Sims	1	700
G. Hinton	4	2150
H. Fu	1	450
H. Taylor	1	400
J. Bain	1	500
L. Nye	2	900
P. Judd	2	900
S. Cann	1	600
T. Goss	2	750

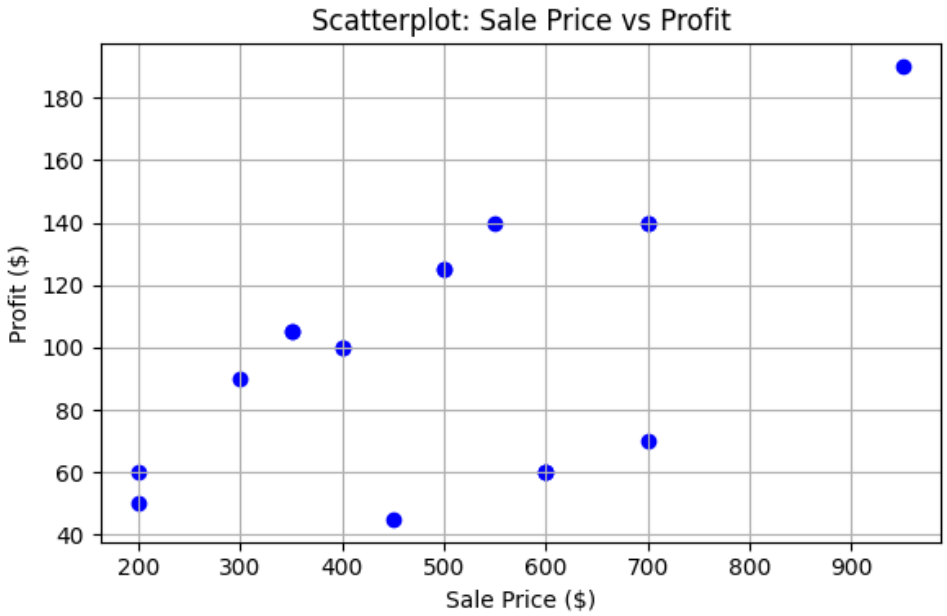
Summary by Store:

Store	Count	Mean_Sale_Price
New York, NY	10	485.0
Washington, DC	8	525.0

Summary by Product Category:

Product Category	Count	Total_Profit
Desktop	5	295
Laptop	3	470
Printer	4	360
Scanner	6	640

(d)

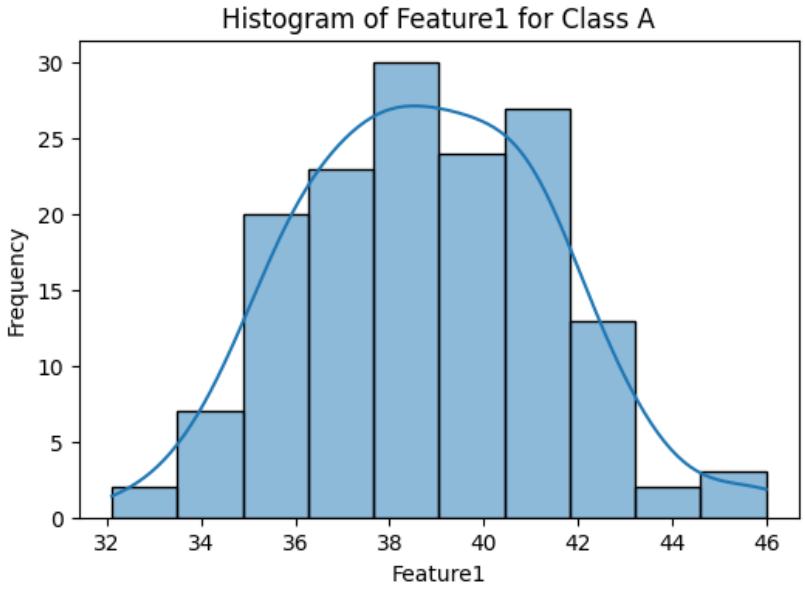


Question 6: Output

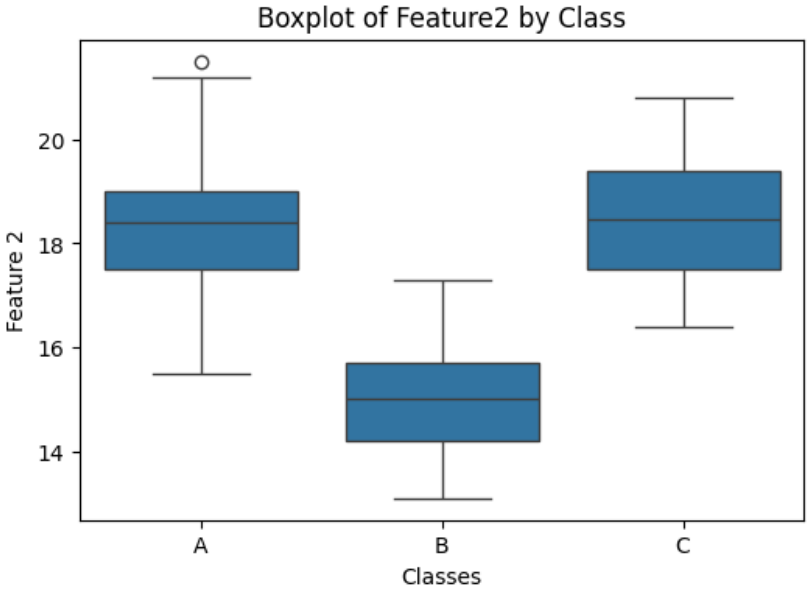
(a) & (b)

```
Frequency of samples per class:
Classes
A    151
B    123
C     68
Name: count, dtype: int64

Descriptive statistics:
       Sample Number   Feature 1   Feature 2   Feature 3    Feature 4
count     342.000000  342.000000  342.000000  342.000000   342.000000
mean      171.500000   43.921930   17.151170  200.915205  4201.754386
std        98.871128    5.459584    1.974793   14.061714   801.954536
min         1.000000   32.100000   13.100000  172.000000  2700.000000
25%        86.250000   39.225000   15.600000  190.000000  3550.000000
50%       171.500000   44.450000   17.300000  197.000000  4050.000000
75%       256.750000   48.500000   18.700000  213.000000  4750.000000
max       342.000000   59.600000   21.500000  231.000000  6300.000000

Interquartile Range (IQR):
Feature 1 IQR: 9.274999999999999
Feature 2 IQR: 3.0999999999999996
Feature 3 IQR: 23.0
Feature 4 IQR: 1200.0
```

(c)


Histogram of Feature1 for Class A
Frequency
Feature1
0
5
10
15
20
25
30
32
34
36
38
40
42
44
46

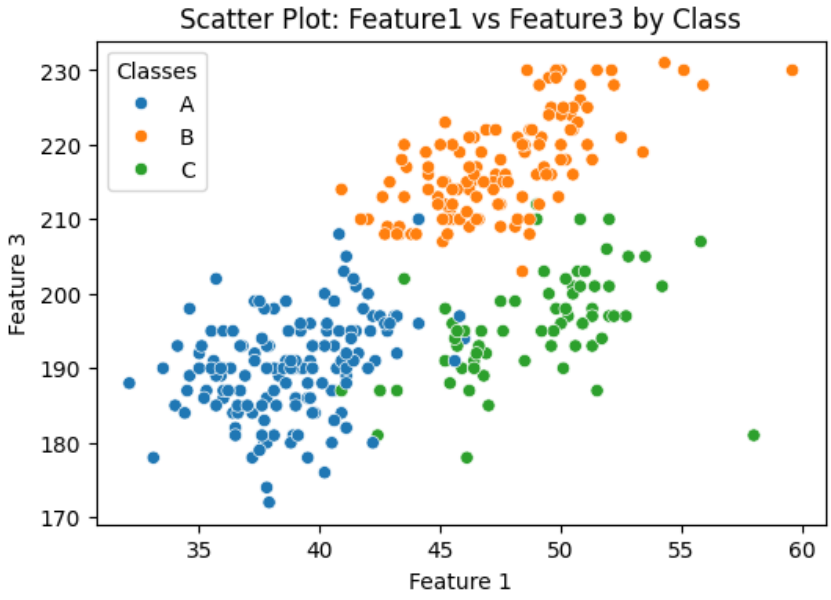
(d)


Boxplot of Feature2 by Class
Feature 2
Classes
14
16
18
20
A
B
C

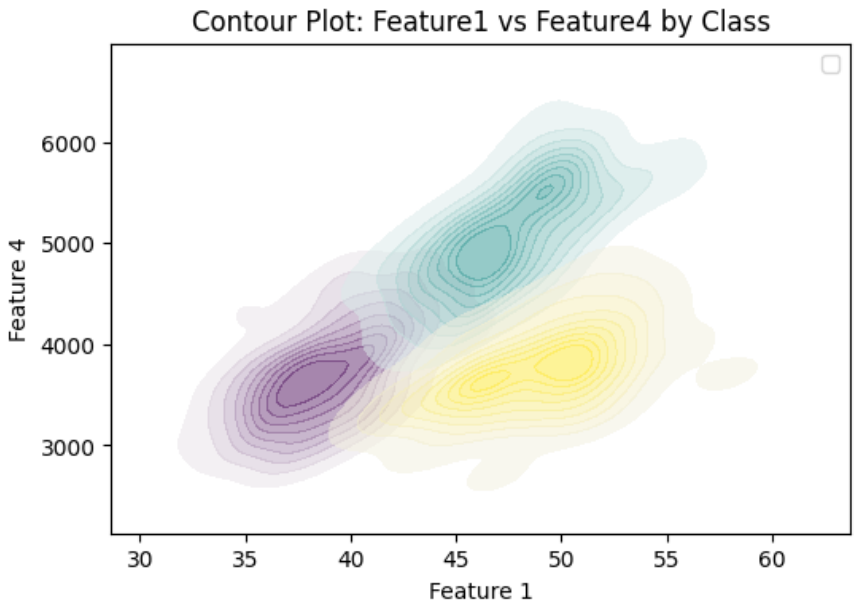
(e)


Violin Plot of Feature3 by Class
Feature 3
230
220
210
200
190
180
170
A
B
C
Classes

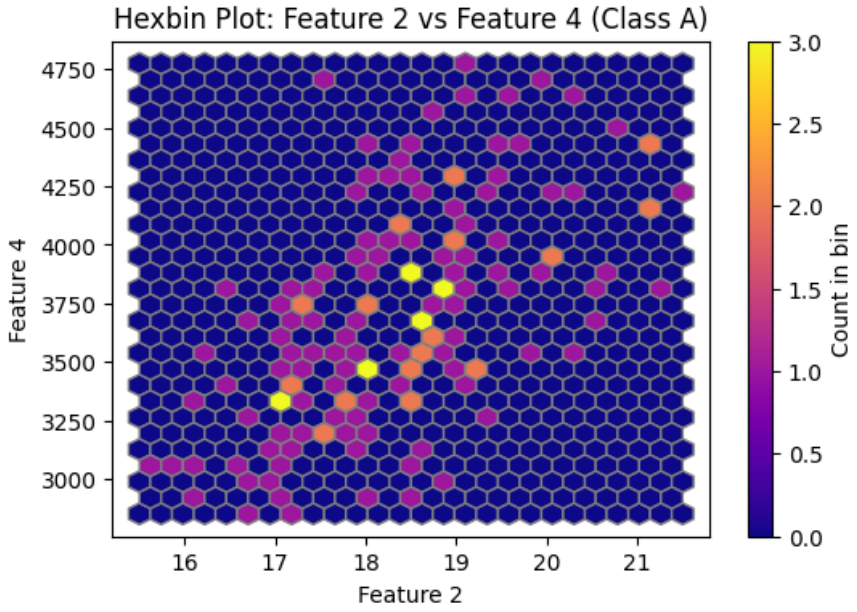
(f)


Scatter Plot: Feature1 vs Feature3 by Class
Classes
A
B
C
Feature 3
230
220
210
200
190
180
170
35
40
45
50
55
60
Feature 1

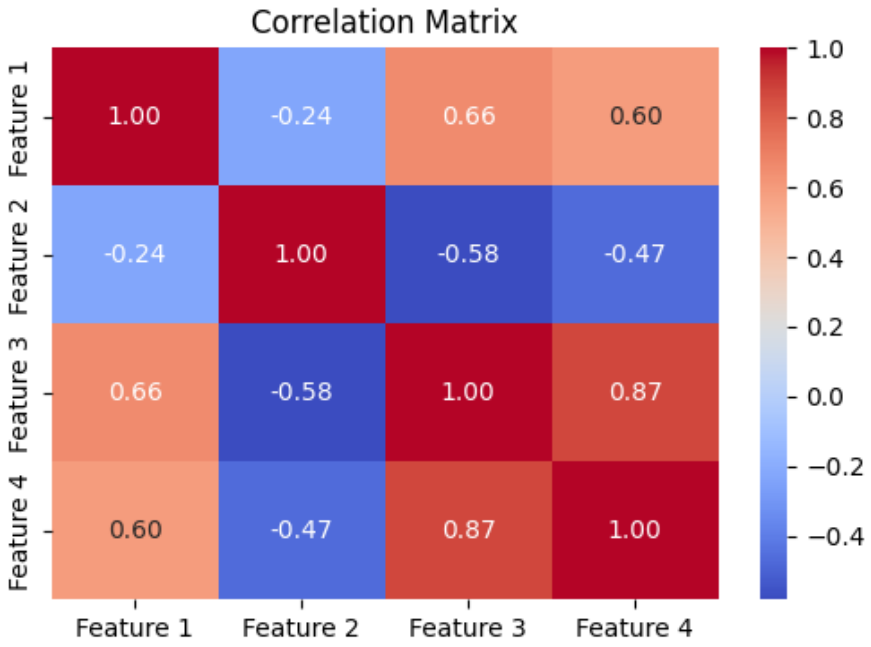
(g)


Contour Plot: Feature1 vs Feature4 by Class
Feature 4
6000
5000
4000
3000
30
35
40
45
50
55
60
Feature 1

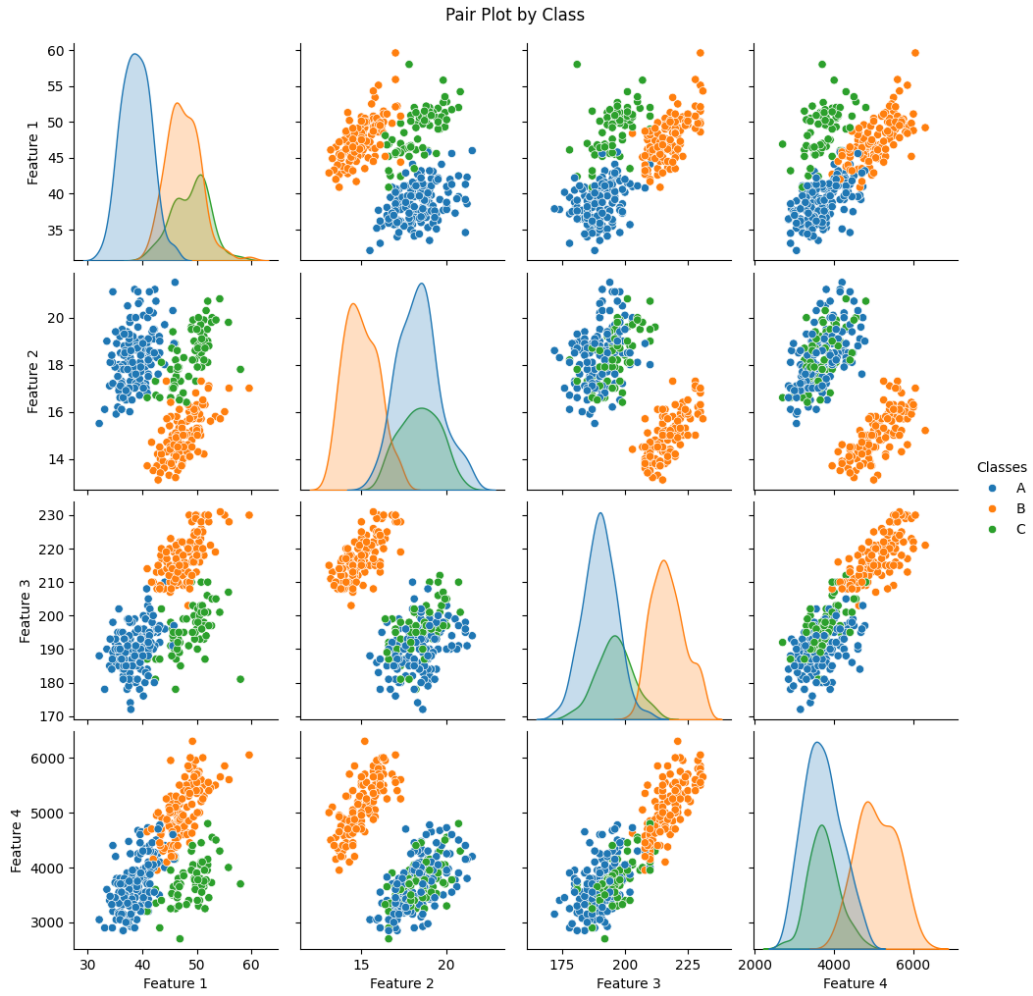
(h)


Hexbin Plot: Feature 2 vs Feature 4 (Class A)
Feature 4
4750
4500
4250
4000
3750
3500
3250
3000
16
17
18
19
20
21
Feature 2
Count in bin
3.0
2.5
2.0
1.5
1.0
0.5
0.0

(i)


Correlation Matrix
Feature 1
Feature 2
Feature 3
Feature 4
1.00
-0.24
0.66
0.60
-0.24
1.00
-0.58
-0.47
0.66
-0.58
1.00
0.87
0.60
-0.47
0.87
1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

(j)


Pair Plot by Class
Feature 1
Feature 2
Feature 3
Feature 4
Classes
A
B
C