Connor Johnson

Assignment 1

1. $u^T * u = 26$

$u * u^T = [16,4,12;4,1,3;12,3,9]$

$v * u = 71$

$u+5 = [9;6;5]$

$A^T = [1,3;2,4;5,6]$

$B * u = [56;19;42]$

$B^{-1} = [1, -5.5, 1.25; 0, -0.5, 0.25; -0.667, 4.333, -1]$

$B+C = [15, 7, 14; 3, -1, 7; 3, 6, 10]$

$B-C = [-1, -5, -4; 1, 5, -1; 5, 10, 2]$

$A * B = [31, 45, 45; 53, 59, 75]$

$B * C = [48, 21, 75; 15, 0, 30; 34, -12, 76]$

$B * A$ = impossible matrix multiplication

2.

1. Min = 0, max = 17, mean = 3.845, stdev = 3.370
2. Min = 0, max = 199, mean = 120.895, stdev = 31.973
3. Min = 0, max = 122, mean = 69.106, stdev = 19.356
4. Min = 0, max = 99, mean = 20.537, stdev = 15.952
5. Min = 0, max = 846, mean = 79.800, stdev = 115.244
6. Min = 0, max = 67.1, mean = 31.993, stdev = 7.884
7. Min = .0780, max = 2.42, mean = 0.472, stdev = 0.331
8. Min = 21, max = 81, mean = 33.241, stdev = 11.760

Class 0

1. Mean = 3.298, stdev = 3.017
2. Mean = 109.980, stdev = 26.141
3. Mean = 68.184, stdev = 18.063
4. Mean = 19.664, stdev 14.890
5. Mean = 68.792, stdev = 98.865
6. Mean = 30.304, stdev 7.690

7. Mean = 0.430, stdev = 0.299
8. Mean = 31.190, stdev = 11.668

Class 1

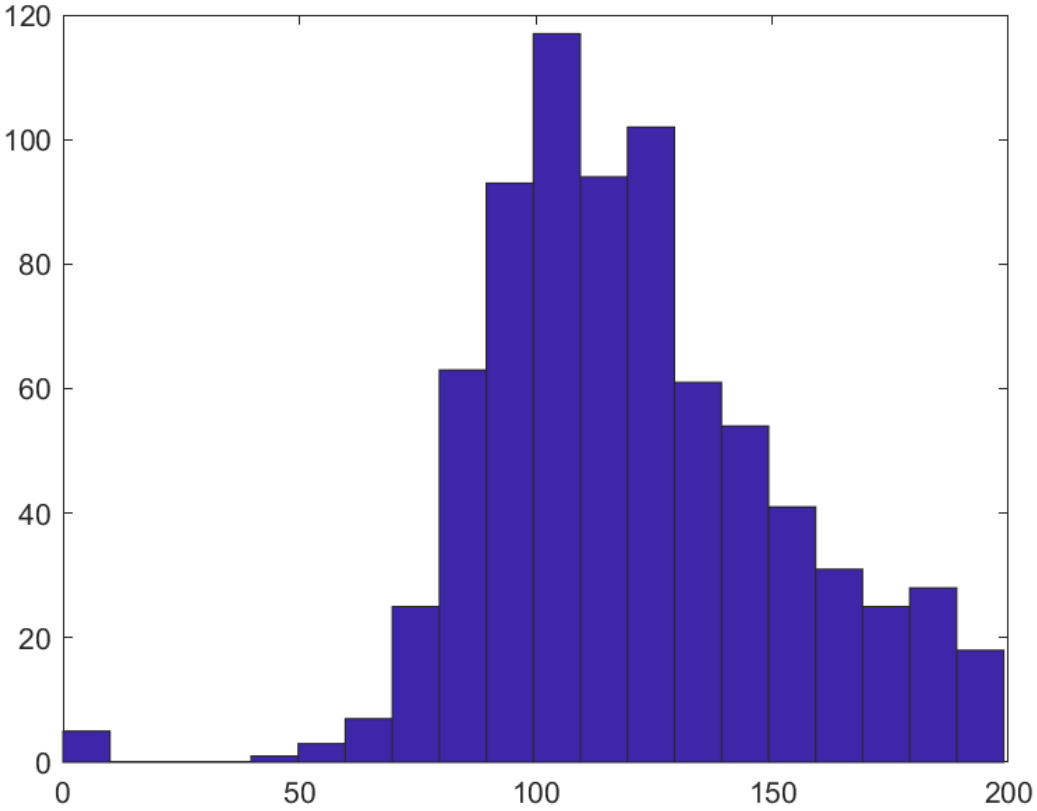
1. Mean = 4.866, stdev = 3.741
2. Mean = 141.258, stdev = 31.940
3. Mean = 70.825, stdev = 21.492
4. Mean = 22.164, stdev = 17.680
5. Mean = 100.336, stdev = 138.689
6. Mean = 35.143, stdev = 7.263
7. Mean = 0.551, stdev = 0.372
8. Mean = 37.067, stdev = 10.968

c) The second attribute or Plasma glucose concentration 2 hours in an oral glucose tolerance test is the most helpful in predicting class. This is because the mean between the 2 classes has the largest difference relative to tis standard deviation. This shows there is a more significant difference in this attribute between the classes so it can be helpful in predicting class.

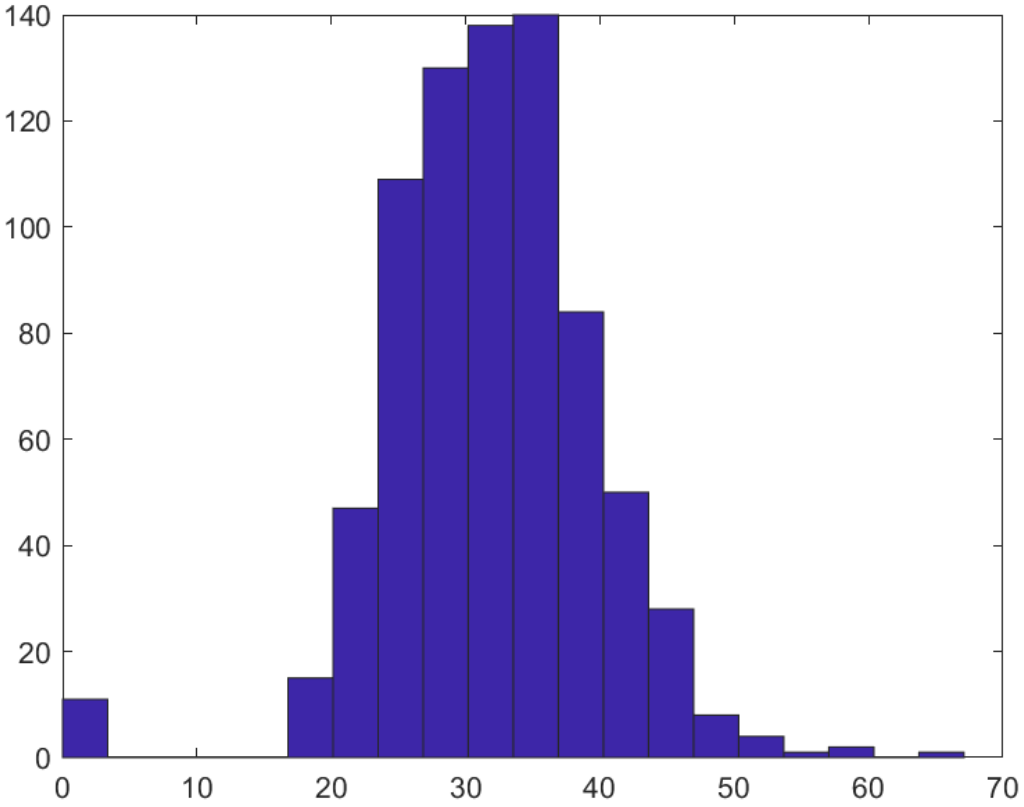
d) The second attribute has the highest correlation with the target attribute at 0.446. So, it is the most helpful in predicting the target class because it is more correlated with it than the other attributes. This means that a change in this attribute is more likely to show a change in the target attribute.

e) The highest mutual correlation is between attributes 1 and 8 with a correlation of 0.5443

g) Attribute 2

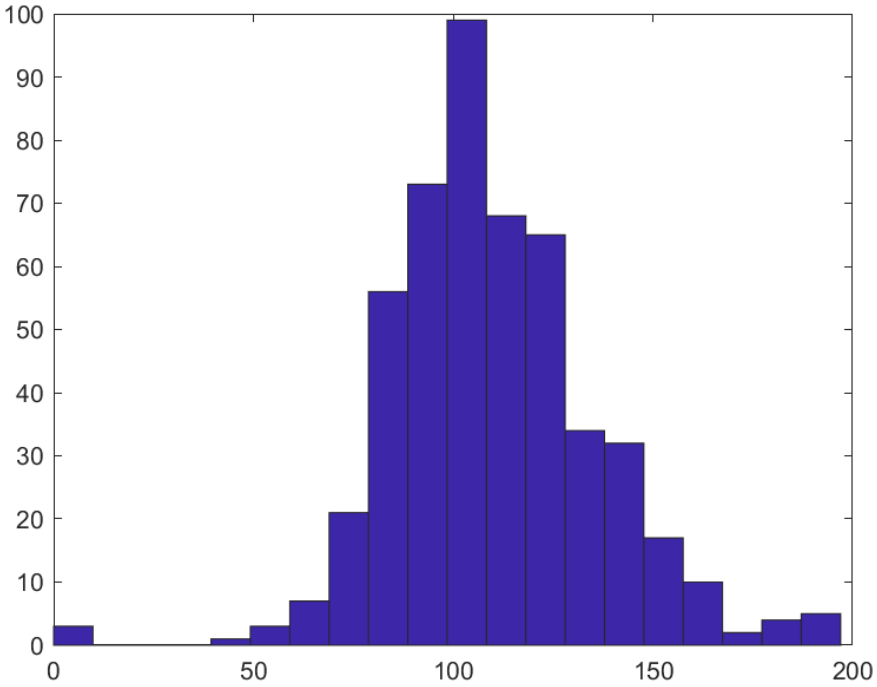

120
100
80
60
40
20
0
0
50
100
150
200

Attribute 6

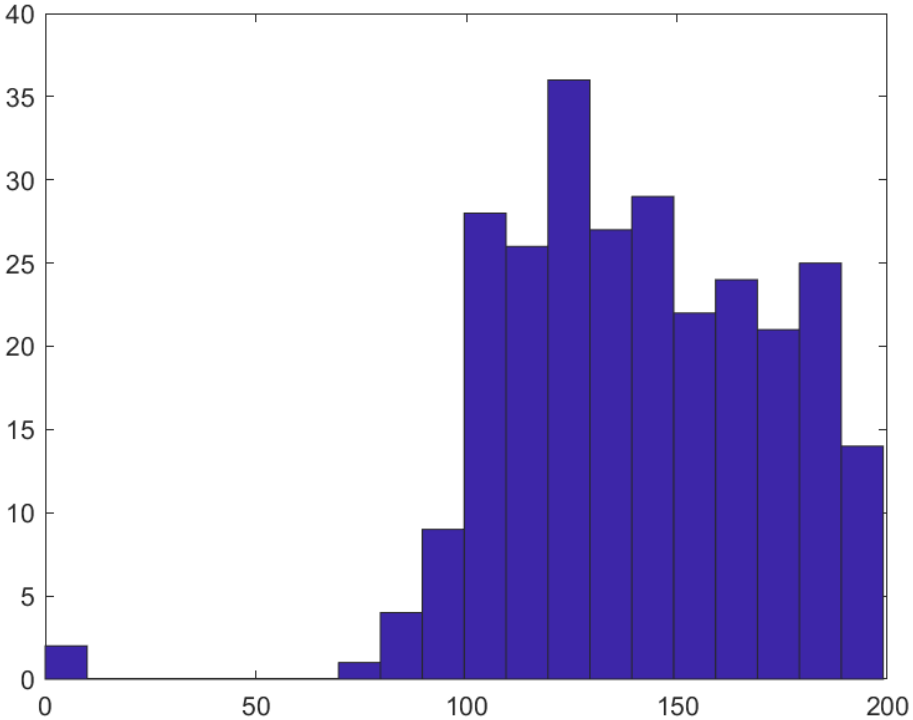

140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
70

These are the 2 attributes with a distribution closest to the normal. Overall Attribute 6 is the most normal because attribute 2 is slightly skewed to the left. They both have outliers toward the left side, but other than this, they're very close to a normal distribution.

h) Attribute 2 Class 0

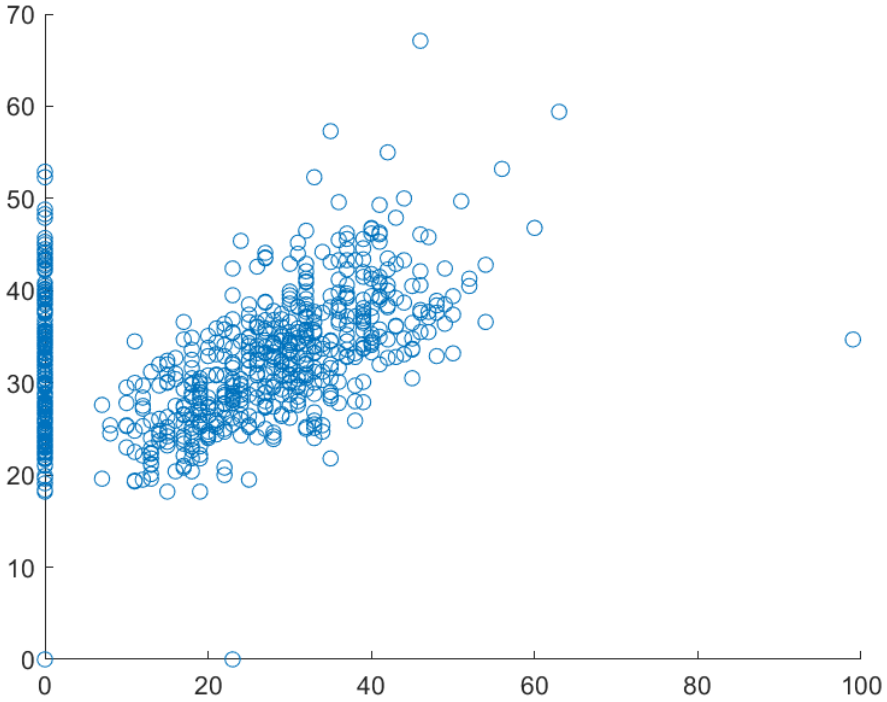



Attribute 2 Class 1

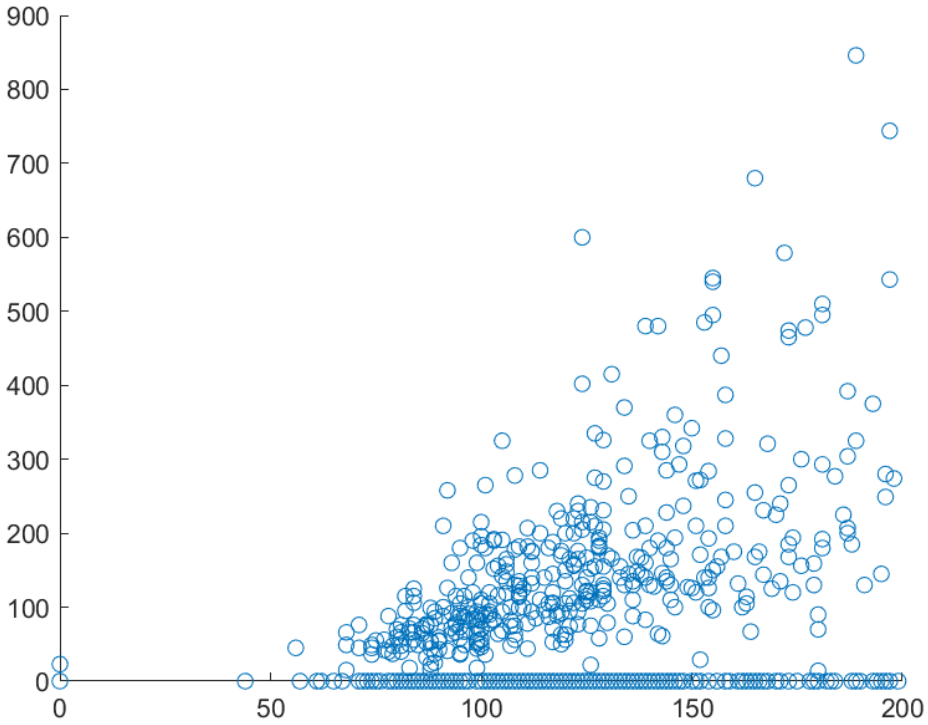


Attribute 2 is best for discriminating the 2 classes because there is a larger shift in the distributions of the classes. The distribution shifts to the left over one standard deviation which is larger than any of the shifts from the other classes.

i) Attributes 4 and 6




Attributes 2 and 5



For 2 independent and random attributes, the scatter plot should have no patter and look like dots randomly scattered on the graph. I found the 2 scatter plots above interesting because they show a clear positive correlation between the attributes other than the some of the zero values.

3.

A) Each unique color is made into an attribute. Since there are 8 colors, the one-hot encoding will include 8 attributes. For each value, the color attribute that it is will be set to one and the rest of them are set to zero.

| | Red | Yellow | Blue | Black | Green | White | Orange | Brown |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 6 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

b) mean = 69.106, stdev = 19.356, [0.150, -0.160, -0.264, -0.160, -1.504]

c) [6, 4, 4, 4, 1]

4.

a) 503, 526, 492, 505, 504, 528, 519, 520, 496, 518, 505, 510, 530, 532, 491, 512, 488, 509, 526, 497

mean = 510.55