1.) Q1

   a. The entropy of this collection is approximately .9991
   b. The information gain compared to entropy is .2294
   c. My guess for the best split would be a1 as a1 produces 0 for the + class
   d. Based on the Gini index, a1 is the best split

   Below is the proof of work

$p: (c0) : 4$

$n: (c1) = 5$

$$E(p,n) = \frac{-p}{p+n} \log \left(\frac{p}{p+n}\right) - \frac{n}{p+n} \log \left(n/_{p+n}\right)$$

$$= \frac{-4}{9} \log_2\left(\frac{4}{9}\right) - 5/9 \log_2\left(5/9\right)$$

$$= .9911$$

information Gain

A1	+	-
T	3	1
F	1	4

$$\frac{4}{9}\left[-\frac{3}{4}\log\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) \log\left(\frac{1}{4}\right)\right]$$

$$+\frac{5}{9}\left[\left(-\frac{1}{5}\right)\log\left(\frac{1}{5}\right) - \left(\frac{4}{5}\right)\log\left(\frac{4}{5}\right)\right]$$

.7616

Compared to entrop

$0.9911 - 0.7616 \approx 0.2294$

best split

a1 produced the best split

What is the best spit between a1 + a2 according to the ginni index?

A1

$$4/9\left[1-(3/4)^2-(1/4)^2\right]+5/9\left[1-(1/5)^2-(4/5)^2\right]$$

$$=.344$$

A2

$$5/9\left[1-(2/5)^2-(3/5)^2\right]+4/9\left[1-(2/4)^2-(2/4)^2\right]$$

$$=0.4889$$

$\boxed{\text{A1 has the better split}}$

1.) Q2

   a. A would be the best option to split
      Below is proof of work

Q2 a.)

Classification error rate

$$1 - \max\left(\left(\frac{50}{100}\right), \left(\frac{50}{100}\right)\right) = \frac{50}{100}$$

A	B	C	+	-
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

	A = T	A = F
+	25	25
-	0	50

$$T = 1 - \max\left(\frac{25}{25}, \frac{0}{25}\right) = \frac{0}{25} = 0$$

$$F = 1 - \max\left(\frac{25}{75}, \frac{0}{75}\right) = \frac{50}{75} = .\overline{333}$$

$\Delta A = .25$

A	B	C	+	-
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

The best attribute to split with is A

	B=T	B=F
+	30	20
-	20	30

$BT = \frac{20}{50}$ $\Delta A = .10$

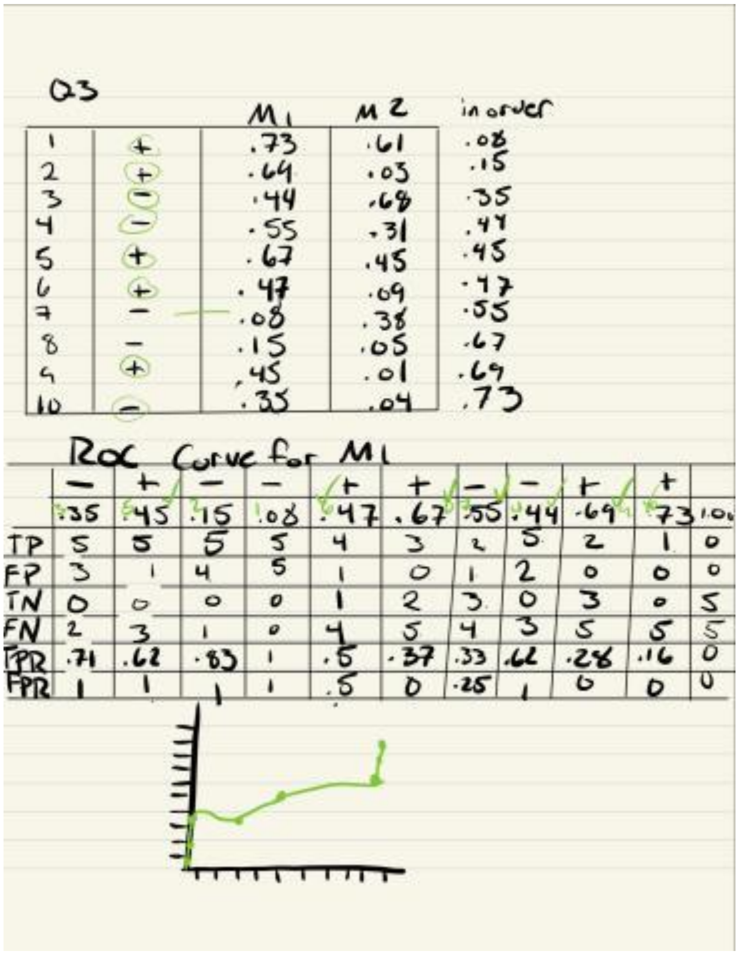
$BF = \frac{20}{50}$

	C=T	C=F
+	25	25
-	25	25

$CT = \frac{25}{50} = 0$

$CF = \frac{25}{50}$

1.) Q3



Q3

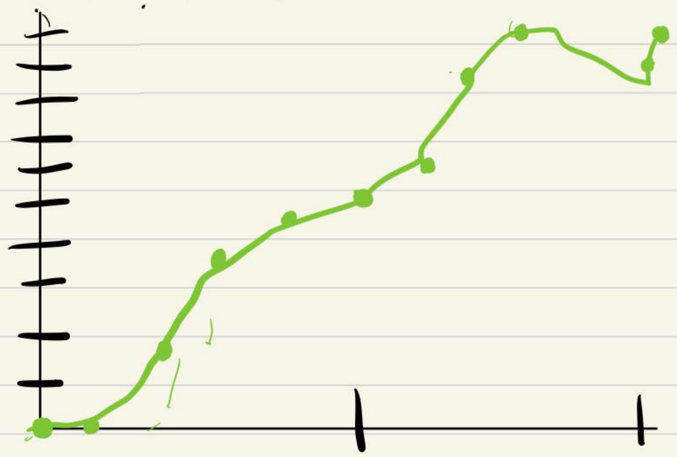
		M1	M2	in order
1	+	.73	.61	.08
2	+	.64	.03	.15
3	−	.44	.68	.35
4	−	.55	.31	.44
5	+	.67	.45	.45
6	+	.47	.09	.47
7	−	.08	.38	.55
8	−	.15	.05	.67
9	+	.45	.01	.69
10	−	.35	.04	.73

ROC Curve for M1

	−	+	−	−	+	+	−	−	+	+	
	.35	.45	.15	.08	.47	.67	.55	.44	.69	.73	1.00
TP	5	5	5	5	4	3	2	5	2	1	0
FP	3	1	4	5	1	0	1	2	0	0	0
TN	0	0	0	0	1	2	3	0	3	0	5
FN	2	3	1	0	4	5	4	3	5	5	5
TPR	.71	.62	.83	1	.5	.37	.33	.62	.28	.16	0
FPR	1	1	1	1	.5	0	.25	1	0	0	0

		M1	M2
1	+	.73	.61 ✓
2	+	.69	.03 ✓
3	−	.44 ✓	.68 ✓
4	−	.55 ✓	.31 ✓
5	+	.67	.45 ✓
6	+	.47 ✓	.09 ✓
7	−	.08 ✓	.38 ✓
8	−	.15 ✓	.05 ✓
9	+	.45 ✓	.01 ✓
10	−	.35 ✓	.04 ✓

	+	+	−	−	+	−	−	+	+	−	
	.01	.03	.04	.05	.09	.31	.38	.45	.61	.68	1.00
TP	5	4	3	3	3	2	2	2	1	0	0
FP	5	5	5	4	3	3	2	1	1	1	0
TN	0	1	0	1	2	3	3	3	4	5	5
FN	0	0	2	2	2	2	3	4	4	4	5
TPR	1	1	.6	.6	.6	.5	.4	.25	.2	0	0
FPR	1	.83	1	.8	.6	.5	.4	.33	.2	.14	0



The graphs have a fairly similar shape to them as compared to my python program