CSC 522 HW 3

Group H29

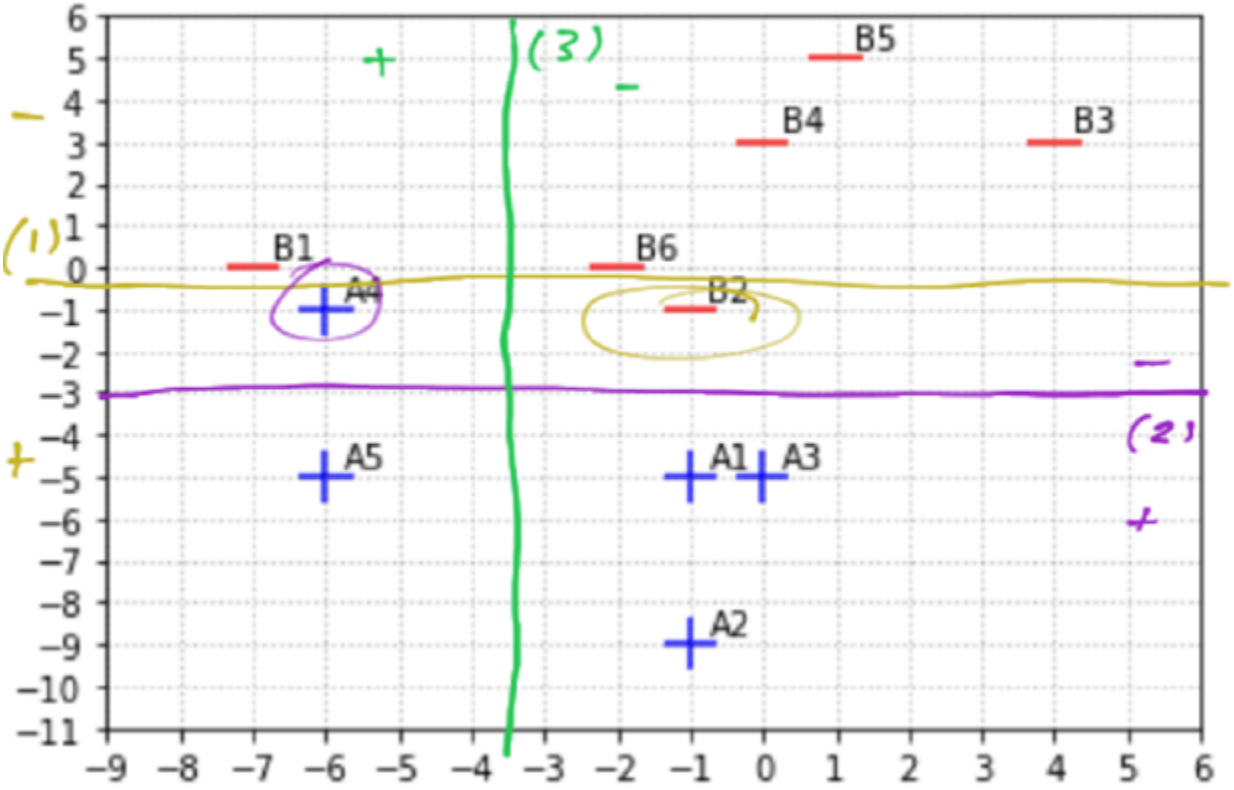
September 29, 2022

Ethan Purnell (efpurne2)
Corey Capooci (cvcapooc)
Jackson Ingraham (jtingrah)
Github Repository: engr-ALDA-Fall2022-H29
`https://github.ncsu.edu/efpurne2/engr-ALDA-Fall2022-H29`

Question 2:

0.1 A, B, D, E

The following decision boundaries were created for parts A, B, D, and E for Figure 1



0.2 B

The following calculations were used to find the highest weight of the first boosting iteration:

$\epsilon = (\text{weights}) * (\text{number of errors}) = (1/11) * 1$
$\alpha = \frac{1}{2} * \ln(\frac{1-\epsilon}{\epsilon}) = 0.1513$
$Z = 1 * (1/11)e^{\alpha} + 10 * (1/11)e^{-\alpha}$

Highest Weight $= \frac{1*(1/11)e^{\alpha}}{Z} = 0.5$

0.3 C

The weighted error for the first decision boundary was calculated using the following:

$\epsilon = (\text{weights}) * (\text{number of errors}) = (0.5) * 1 = 0.5$

0.4 E

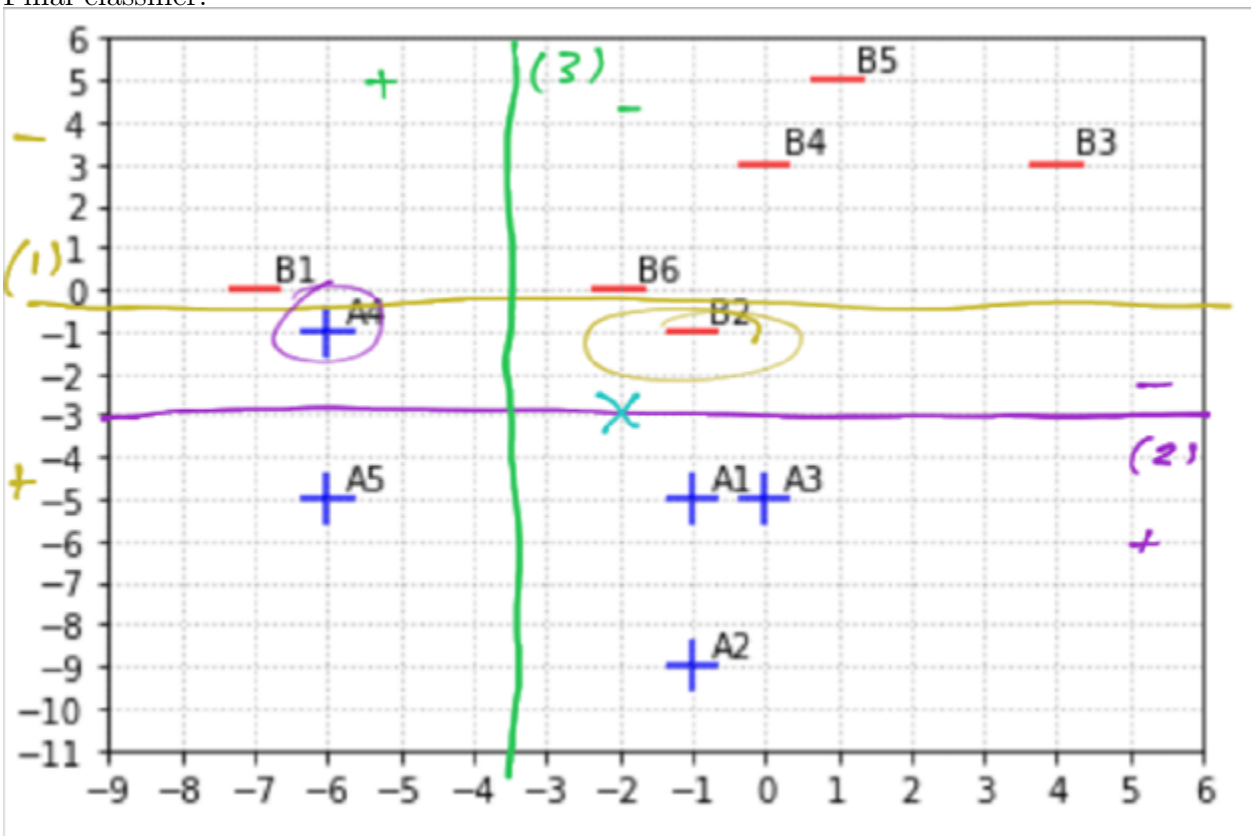
Since the point A4 was the only error point in the decision boundary 2, it's weight is the weighted error of decision boundary 2. It was calculated using the following:

$\epsilon = 0.05 * 1 = 0.05$
$\alpha = \frac{1}{2} * \ln(\frac{1-\epsilon}{\epsilon}) = 1.4722$
$Z = 1 * 0.5e^{-\alpha} + 1 * 0.05e^{\alpha} + 9 * 0.05e^{-\alpha}$

Weight of A4 = Weighted ϵ of decision boundary $2 = 1 * 0.05e^{\alpha}/Z = 0.5$

0.5 F

Final classifier:



Since the added point lands on the decision boundary of 2, the point could be classified as either '-' or '+'. Using the piazza response about tie-breaking points on the boundary line, this point would default to being defined as a '+'.

Question 3:

3a.

Naive Bayes

Fold 0
For the test data,

For ID1,
$P(A|No) = \frac{3}{7} * \frac{3}{7} * \frac{2}{7} * \frac{1}{7} * \frac{5}{7} = 0.007$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose Yes.

For ID3,
$P(A|No) = \frac{4}{7} * \frac{4}{7} * \frac{5}{7} * \frac{6}{7} * \frac{5}{7} = 0.143$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose No.

For ID5,
$P(A|No) = \frac{3}{7} * \frac{4}{7} * \frac{2}{7} * \frac{6}{7} * \frac{5}{7} = 0.0428$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose No.

For ID7,
$P(A|No) = \frac{3}{7} * \frac{3}{7} * \frac{2}{7} * \frac{1}{7} * \frac{5}{7} = 0.00535$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose Yes.

For ID9,
$P(A|No) = \frac{4}{7} * \frac{4}{7} * \frac{2}{7} * \frac{6}{7} * \frac{5}{7} = 0.0571$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose No.

For ID11,
$P(A|No) = \frac{4}{7} * \frac{4}{7} * \frac{2}{7} * \frac{1}{7} * \frac{5}{7} = 0.00952$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose Yes.

For ID13,
$P(A|No) = \frac{4}{7} * \frac{3}{7} * \frac{2}{7} * \frac{1}{7} * \frac{5}{7} = 0.00714$
$P(A|Yes) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * \frac{2}{7} = 0.0179$
Choose Yes.

Fold 1

For ID2,
$P(A|No) = \frac{3}{7} * \frac{5}{7} * \frac{2}{7} * \frac{4}{7} * \frac{5}{7} = 0.0357$
$P(A|Yes) = \frac{3}{4} * \frac{2}{4} * \frac{1}{4} * \frac{1}{4} * \frac{2}{7} = 0.0067$
Choose No.

For ID4,
$P(A|No) = \frac{4}{7} * \frac{5}{7} * \frac{2}{7} * \frac{4}{7} * \frac{5}{7} = 0.0476$
$P(A|Yes) = \frac{1}{4} * \frac{2}{4} * \frac{1}{4} * \frac{1}{4} * \frac{2}{7} = 0.0022$
Choose No.

For ID6,
$P(A|No) = \frac{4}{7} * \frac{5}{7} * \frac{2}{7} * \frac{4}{7} * \frac{5}{7} = 0.0476$
$P(A|Yes) = \frac{1}{4} * \frac{2}{4} * \frac{1}{4} * \frac{1}{4} * \frac{2}{7} = 0.0022$
Choose No.

For ID8,
$P(A|No) = \frac{3}{7} * \frac{2}{7} * \frac{2}{7} * \frac{4}{7} * \frac{5}{7} = 0.0143$
$P(A|Yes) = \frac{3}{4} * \frac{2}{4} * \frac{1}{4} * \frac{1}{4} * \frac{2}{7} = 0.0067$
Choose No.

For ID10,
$P(A|No) = \frac{3}{7} * \frac{2}{7} * \frac{5}{7} * \frac{4}{7} * \frac{5}{7} = 0.0357$
$P(A|Yes) = \frac{3}{4} * \frac{2}{4} * \frac{3}{4} * \frac{1}{4} * \frac{2}{7} = 0.020$
Choose No.

For ID12,
$P(A|No) = \frac{4}{7} * \frac{5}{7} * \frac{5}{7} * \frac{4}{7} * \frac{5}{7} = 0.119$
$P(A|Yes) = \frac{1}{4} * \frac{2}{4} * \frac{3}{4} * \frac{1}{4} * \frac{2}{7} = 0.0067$
Choose No.

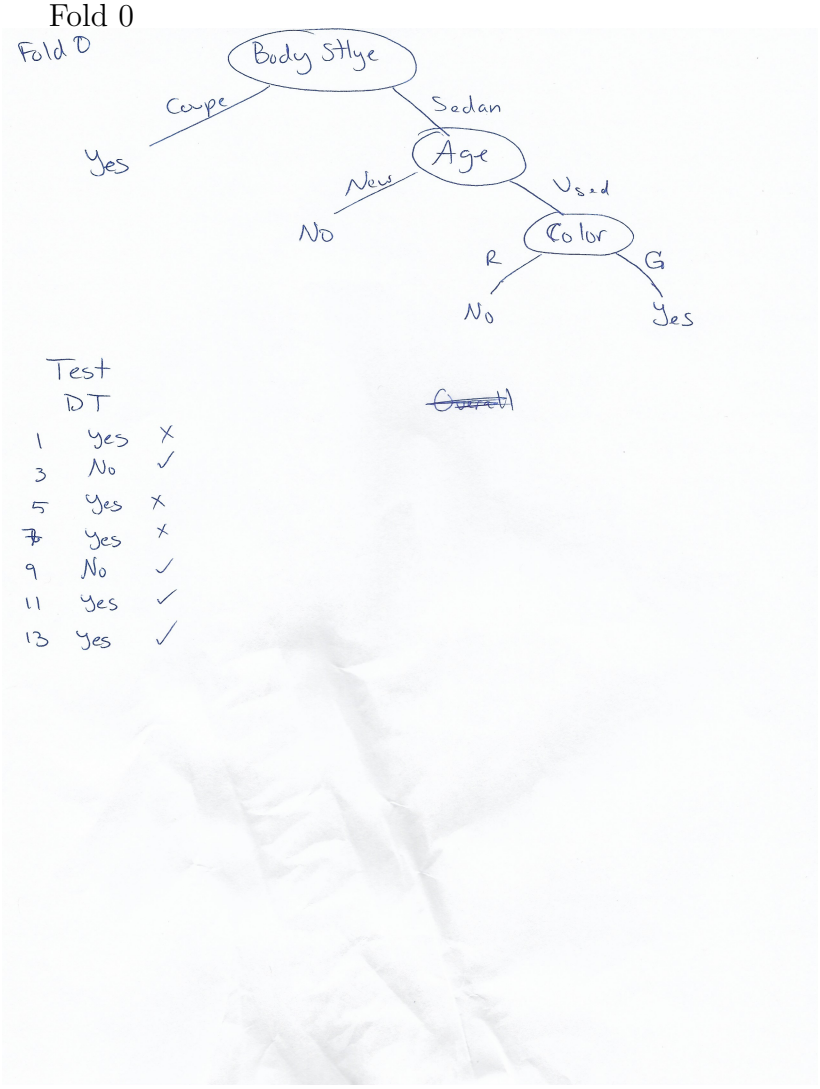
For ID14,
$P(A|No) = \frac{3}{7} * \frac{2}{7} * \frac{2}{7} * \frac{3}{7} * \frac{5}{7} = 0.0107$
$P(A|Yes) = \frac{3}{4} * \frac{2}{4} * \frac{1}{4} * \frac{3}{4} * \frac{2}{7} = 0.020$
Choose Yes.

Result
The 2-fold accuracy is $\frac{11}{14}$

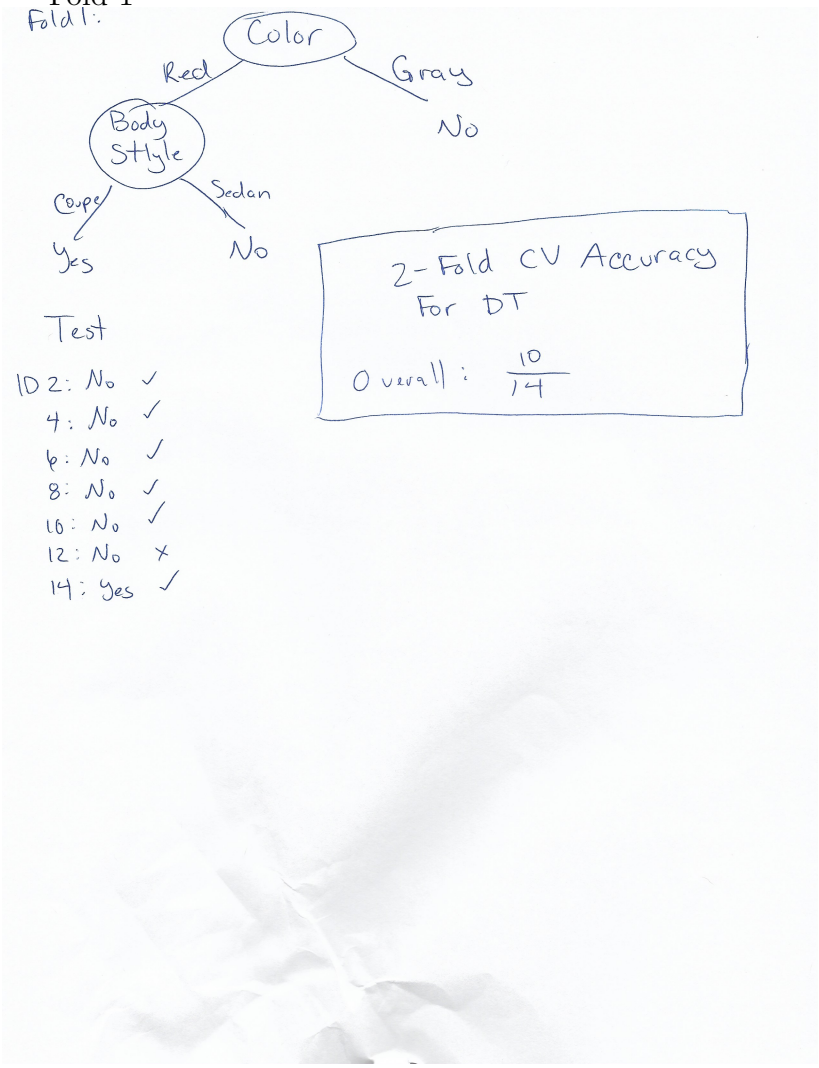
Decision Tree

Additional information and calculations for the decision tree is in the appendix. To remain consistent with what was requested in problem 3a, only the completed DT models are shown below.

Fold 0



Fold 1


Fold 1:
Color
Red
Gray
No
Body Style
Coupe
Sedan
Yes
No
Test
ID 2: No ✓
4: No ✓
6: No ✓
8: No ✓
10: No ✓
12: No ×
14: Yes ✓
2-Fold CV Accuracy For DT
Overall: 10/14

Result

The 2-fold accuracy is $\frac{10}{14}$

3b.

Based on the 2-fold CV accuracy, the NB classifier would be chosen.
The classifier is.

$$P(Color = Red|No) = \frac{6}{12}$$

$$P(Color = Gray|No) = \frac{6}{12}$$

$$P(Transmission = Manual|No) = \frac{8}{12}$$

$$P(Transmission = Automatic|No) = \frac{4}{12}$$

$$P(Age = New|No) = \frac{6}{12}$$

$$P(Age = Used|No) = \frac{6}{12}$$

$$P(BodyStyle = Sedan|No) = \frac{9}{12}$$

$$P(BodyStyle = Coupe|No) = \frac{3}{12}$$

$$P(Color = Red|Yes) = \frac{4}{6}$$

$$P(Color = Gray|Yes) = \frac{2}{6}$$

$$P(Transmission = Manual|Yes) = \frac{3}{6}$$

$$P(Transmission = Automatic|Yes) = \frac{3}{6}$$

$$P(Age = New|Yes) = \frac{2}{6}$$

$$P(Age = Used|Yes) = \frac{4}{6}$$

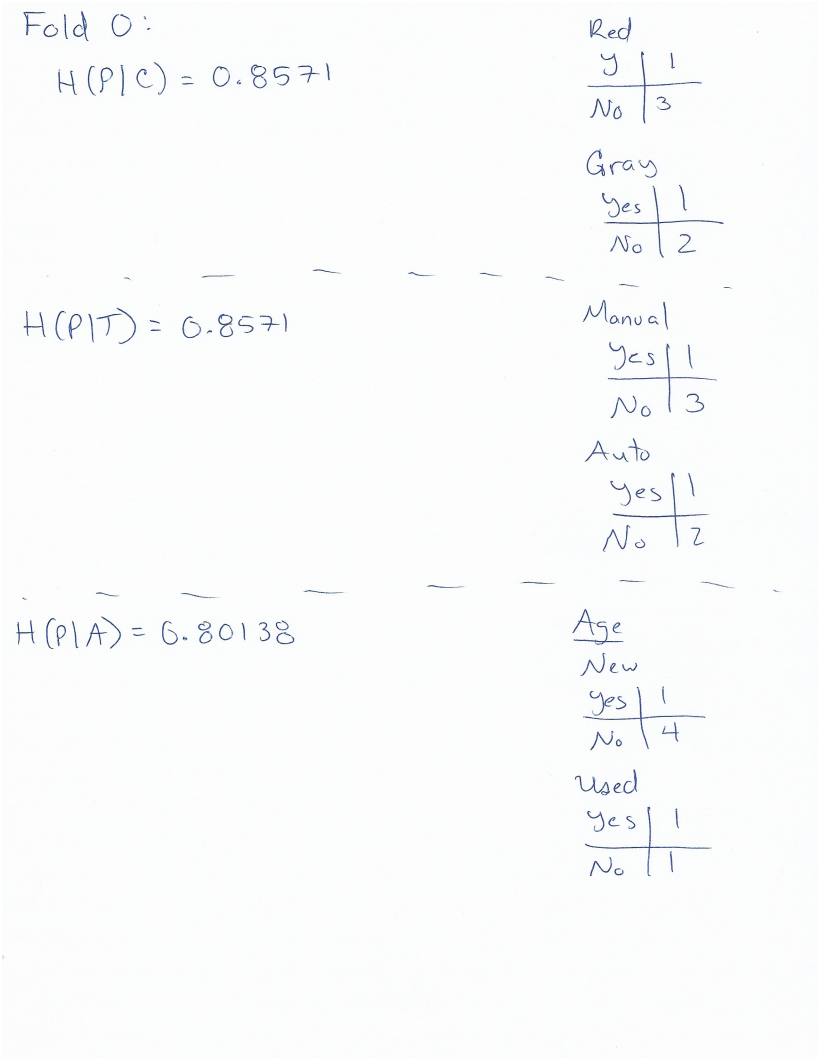
$$P(BodyStyle = Sedan|Yes) = \frac{2}{6}$$

$$P(BodyStyle = Coupe|Yes) = \frac{4}{6}$$

Appendix

3a

Fold 0



$H(PIB) = 0.557$

BS

Sedan

Yes	1
No	5

Coupe

Yes	1
No	0

Choose Body Style

Given Body Style is Sedan

$H(P|C) = 0.459$

Red

Yes	0
No	3

Gray

Yes	1
No	2

$H(P|T) =$ ~~0.5409~~ 0.5409

Manual

Yes	1
No	3

Auto

Yes	0
No	2

$H(P|A) = 0.333$

New

Yes	0
No	2

Used

Yes	1
No	1

Choose Age

Given Body is Sedan, Age is Used

$H(P|T) = 0$

Manual

Yes	1
No	0

Auto

Yes	0
No	1

$H(P|C) = 0$

Red

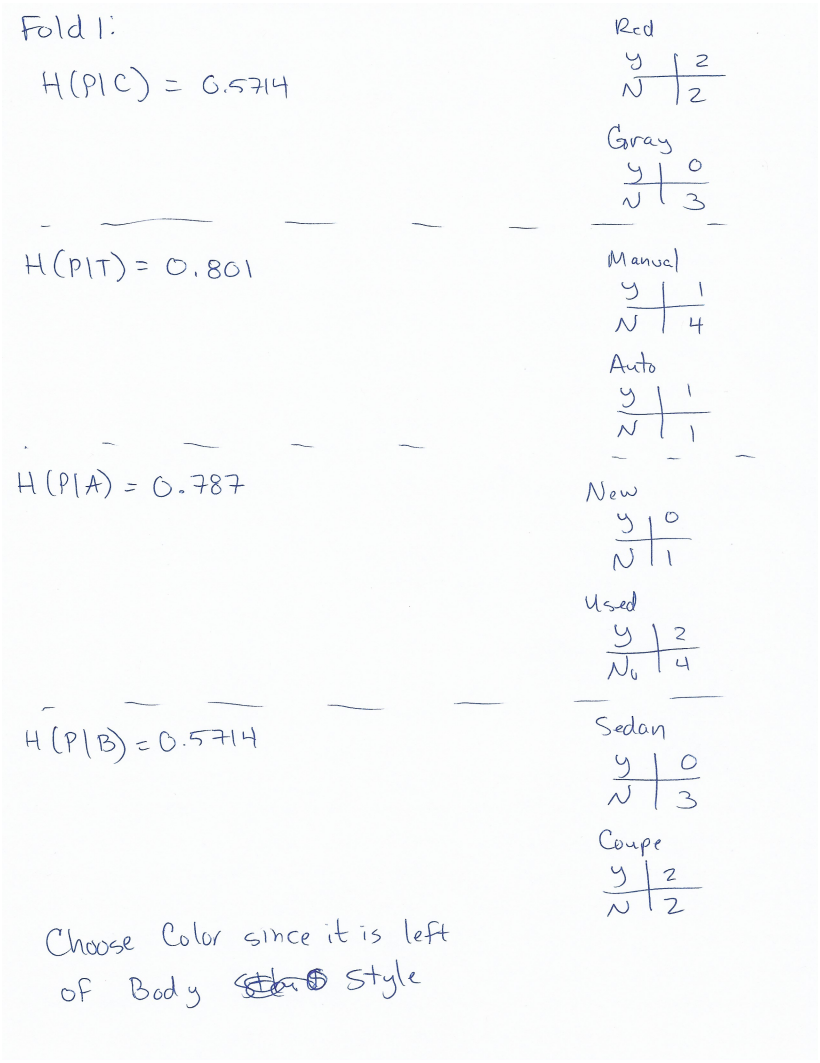
Yes	0
No	1

Gray

Yes	1
No	0

Choose Color since it is leftmost

Fold 1


Fold 1:
H(P|C) = 0.5714
Red
y | 2
N | 2
Gray
y | 0
N | 3
H(P|T) = 0.801
Manual
y | 1
N | 4
Auto
y | 1
N | 1
H(P|A) = 0.787
New
y | 0
N | 1
Used
y | 2
No | 4
H(P|B) = 0.5714
Sedan
y | 0
N | 3
Coupe
y | 2
N | 2
Choose Color since it is left
of Body Style

Given Color is Red

Auto	
y	1
N	0

New	
y	0
N	1

Body Style

Sedan	
y	0
N	2

Man	
y	1
N	2

Used	
y	2
N	1

Coupe	
y	2
N	0

~~Since Body Style is~~

Since $H(P|B) = 0$ we choose body style.