CS548 homework

Cate Dunham

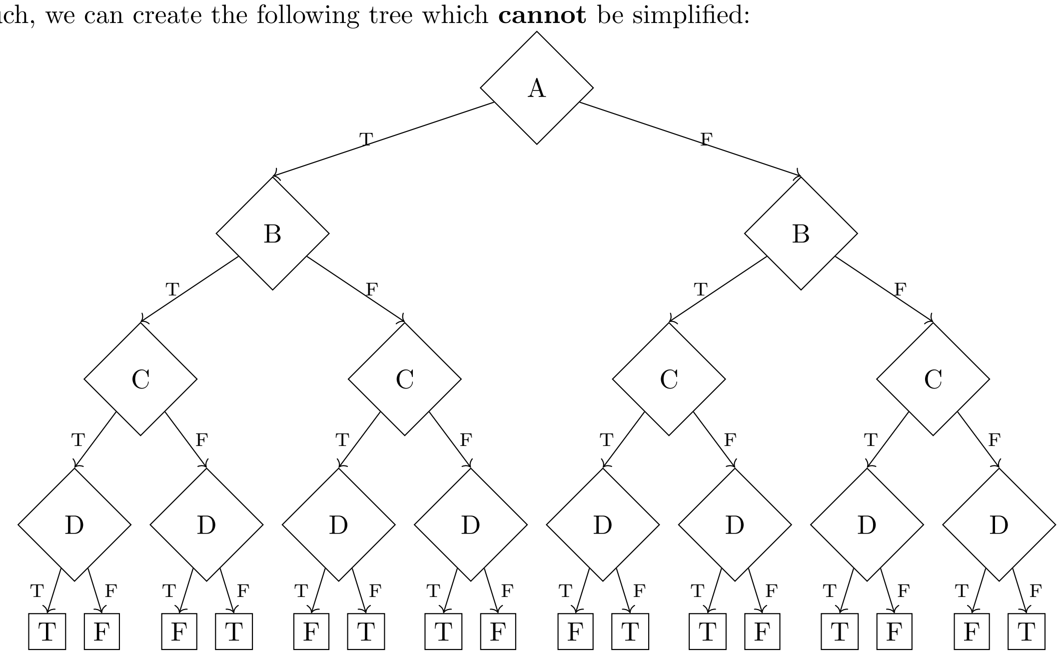
March 3, 2025

Exercise 3.1

10 points

Draw the full decision tree for the parity function of four Boolean attributes, A, B, C, D. Is it possible to simplify the tree?

We know from the textbook that in a parity function the value is 0 (1) when there is an odd (even) number of Boolean attributes with the value *True*. As such, we can create the following tree which **cannot** be simplified:



Exercise 3.2

21 points

Consider the training examples shown in Table 3.5 for a binary classification problem.

Table 3.1. Data set for Exercise 2.

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

a) (3 points) Compute the Gini index for the overall collection of training examples. Since half of the examples fall in each class, the Gini index is:

$$Gini = 1 - (.5^2 + .5^2) = \mathbf{.5}$$

b) (3 points) Compute the Gini index for the Customer ID attribute.

$$Gini_1 = 1 - \frac{1}{1}^2 = 0$$

$$Gini_2 = 1 - \frac{1}{1}^2 = 0$$

$$\forall_i \in \{1, 2, \ldots, n\}, Gini_i = 0$$

$$Gini = .1 \cdot Gini_1 + .1 \cdot Gini_2 + \ldots + .1 \cdot Gini_n = \mathbf{0}$$

c) (3 points) Compute the Gini index for the Gender attribute.

$$Gini_M = 1 - (.6^2 + .4^2) = .48$$

$$Gini_F = 1 - (.4^2 + .6^2) = .48$$

$$Gini = .5 \cdot Gini_M + .5 \cdot Gini_F = \mathbf{.48}$$

d) (3 points) Compute the Gini index for the Car Type attribute using multiway split.

$$Gini_{Fam} = 1 - (\frac{1}{4}^2 + \frac{3}{4}^2) = .375$$

$$Gini_S = 1 - 1^2 = 0$$

$$Gini_L = 1 - (\frac{1}{8}^2 + \frac{7}{8}^2) = .21875$$

$$Gini = .2 \cdot .375 + .4 \cdot 0 + .4 \cdot .21875 = \mathbf{.1625}$$

e) (3 points) Compute the Gini index for the Shirt Size attribute using multiway split.

$$Gini_{SM} = 1 - (\frac{3}{5}^2 + \frac{2}{5}^2) = .48$$

$$Gini_M = 1 - (\frac{3}{7}^2 + \frac{4}{7}^2) = .4898$$

$$Gini_{LG} = 1 - (\frac{2}{4}^2 + \frac{2}{4}^2) = .5$$

$$Gini_{XL} = 1 - (\frac{2}{4}^2 + \frac{2}{4}^2) = .5$$

$$Gini = .25 \cdot .48 + .35 \cdot .4898 + .2 \cdot .5 + .2 \cdot .5 = \mathbf{.49143}$$

f) (3 points) Which attribute is better, Gender, Car Type, or Shirt Size?

With the lowest gini of the three, .1625, Car Type is the best attribute.

g) (3 points) Explain why Customer ID should not be used as the attribute test condition even though it has the lowest Gini.

Each customer has their own, unique, Customer ID, making it useless for prediction.

Exercise 3.3

18 points

Consider the training examples shown in Table 3.6 for a binary classification problem.

a) (3 points) What is the entropy of this collection of training examples with respect to the class attribute?

$$P(+) = 4/9, P(-) = 5/9$$

$$Entropy = -4/9 \log_2 4/9 - 5/9 \log_2 5/9 = .52 + .4711 = \mathbf{.9911}$$

Instance	a_1	a_2	a_3	Target Class
1	T	T	1.0	$+$
2	T	T	6.0	$+$
3	T	F	5.0	$-$
4	F	F	4.0	$+$
5	F	T	7.0	$-$
6	F	T	3.0	$-$
7	F	F	8.0	$-$
8	T	F	7.0	$+$
9	F	T	5.0	$-$

Table 3.6

a_1	+	-
T	3	1
F	1	4

a_1 Contingency Table

b) (3 points) What is the information gains of a_1 and a_2 relative to these training examples?

$$E_{a_{1_T}} = \frac{3}{4} log_2 \frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = -.31125 - .5 = -.81125$$
$$E_{a_{1_F}} = \frac{1}{5} log_2 \frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} = -.4644 - .25752 = .72192$$

$$\Delta a_1 = E - \frac{4}{9}E_{a_{1_T}} - \frac{5}{9}E_{a_{1_F}} = .9911 - .76163 = \mathbf{.22947}$$

a_2	+	-
T	2	3
F	2	2

a_2 Contingency Table

$$E_{a_{2_T}} = \frac{2}{5} log_2 \frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = -.5288 - .4422 = -.971$$
$$E_{a_{2_F}} = \frac{2}{4} log_2 \frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = -.5 - .5 = -1$$

$$\Delta a_2 = E - \frac{5}{9}E_{a_{2_T}} - \frac{4}{9}E_{a_{2_F}} = .9911 - .539444 - .444445 = \mathbf{.00721}$$

c) (3 points) For a_3, which is a continuous attribute, compute the information gain for every possible split.

Split at 2:

$$\Delta 2 = E - 1/9(\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1}) - 8/9(\frac{3}{8}\log_2\frac{3}{8} - \frac{5}{8}\log_2\frac{5}{8}) = \mathbf{.143}$$

Split at 3.5:

$$\Delta 3.5 = E - 2/9(\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) - 7/9(\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7}) = \mathbf{.0026}$$

Split at 3.5:

$$\Delta 3.5 = E - 2/9(\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) - 7/9(\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7}) = \mathbf{.0026}$$

Split at 4.5:

$$\Delta 4.5 = E - 3/9(\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}) - 6/9(\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}) = \mathbf{.0728}$$

Split at 5.5:

$$\Delta 5.5 = E - 5/9(\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}) - 4/9(\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}) = \mathbf{.0072}$$

Split at 6.5:

$$\Delta 6.5 = E - 6/9(\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6}) - 3/9(\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}) = \mathbf{.0183}$$

Split at 7.5:

$$\Delta 7.5 = E - 8/9(\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8}) - 1/9(\frac{0}{1}\log_2\frac{0}{1} - \frac{1}{1}\log_2\frac{1}{1}) = \mathbf{.1022}$$

d) (3 points) What is the best split (among a1, a2, and a3) according to the information gain?

The best split for information gain belongs to a_1.

e) (3 points) What is the best split (between a1 and a2) according to the misclassification error rate?

$$CE_{a_{1_T}} = 1 - \max\left[\frac{3}{4}, \frac{1}{4}\right] = \frac{1}{4}$$

$$CE_{a_{1_F}} = 1 - \max\left[\frac{1}{5}, \frac{4}{5}\right] = \frac{1}{5}$$

$$\Delta a_1 = \frac{1}{4}\cdot\frac{4}{9} + \frac{1}{5}\cdot\frac{5}{9} = \mathbf{\frac{2}{9}}$$

$$CE_{a_{2_T}} = 1 - \max\left[\frac{2}{5}, \frac{3}{5}\right] = \frac{2}{5}$$

$$CE_{a_{1_F}} = 1 - \max\left[\frac{2}{4}, \frac{2}{4}\right] = \frac{2}{4}$$

$$\Delta a_2 = \frac{2}{5} \cdot \frac{5}{9} + \frac{2}{4} \cdot \frac{4}{9} = \mathbf{\frac{4}{9}}$$

Using misclassification error we would select a_1.

f) (3 points) What is the best split (between a1 and a2) according to the Gini index?

$$G_{a_1} = \frac{4}{9}\left[1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right] + \frac{5}{9}\left[1 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2\right] = \mathbf{.3444}$$

$$G_{a_2} = \frac{5}{9}\left[1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2\right] + \frac{4}{9}\left[1 - \left(\frac{2}{4}\right)^2 - \left(\frac{3}{4}\right)^2\right] = \mathbf{.4889}$$

Based on Gini we would pick a_1

Exercise 3.5

9 points

Consider the following data set for a binary classification problem.

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	−
T	T	+
F	F	−
F	F	−
F	F	−
T	T	−
T	F	−

a) (3 points) Calculate the information gain in terms of Entropy when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

A	+	-
T	4	3
F	0	3

B	+	-
T	3	1
F	1	5

Entropy before split:

$$\begin{aligned} \text{E} &= -(4/10)\log_2(4/10) - (6/10)\log_2(6/10) && (1) \\ &= -.4 * -1.322 - .6 * -.737 && (2) \\ &= \mathbf{.971} && (3) \end{aligned}$$

Split on A:

$$E_{A_T} = -\frac{4}{7}\log_2\frac{4}{7} - \frac{3}{7}\log_2\frac{3}{7} = .9852$$

$$E_{A_F} = -\frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3} = 0$$

$$\Delta A = E - \frac{7}{10}E_{A_T} - \frac{3}{10}E_{A_F} = \mathbf{.2813}$$

Split on B:

$$E_{B_T} = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = .8113$$

$$E_{B_F} = -\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6} = .65$$

$$\Delta B = E - \frac{4}{10}E_T - \frac{6}{10}E_F = \mathbf{.2565}$$

Using Entropy, A will be chosen for the first split.

b) (3 points) Calculate the gain in the Gini index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

Gini before split:

$$G = 1 - (\frac{4}{10})^2 - (\frac{6}{10})^2 = .48$$

Split on A:

$$G_{A_T} = 1 - (\frac{4}{7})^2 - (\frac{3}{7})^2 = .4898$$

$$G_{A_F} = 1 - (\frac{0}{3})^2 - (\frac{3}{3})^2 = 0$$

$$\Delta A = G - \frac{7}{10}G_{A_T} - \frac{3}{10}G_{A_F} = \mathbf{.13714}$$

Split on B:

$$G_{B_T} = 1 - (\frac{3}{4})^2 - (\frac{1}{4})^2 = .375$$

$$G_{B_F} = 1 - (\frac{1}{6})^2 - (\frac{5}{6})^2 = .278$$

$$\Delta B = G - \frac{4}{10}G_{B_T} - \frac{6}{10}G_{B_F} = \mathbf{.1633}$$

Using Gini we would split on B

c) (3 points) Figure 3.11 shows that entropy and the Gini index are both monotonically increasing on the range [0, 0.5] and they are both monotonically decreasing on the range [0.5, 1]. Is it possible that information gain and the gain in the Gini index favor different attributes? Explain.

Yes this is definitely possible. If, for example, one attribute creates splits with pure nodes it might be favored by an Entropy-based system even if that feature is not the best feature to select for reducing misclassification.

Exercise 3.8

15 points

The following table summarizes a data set with three different attributes A, B, C and two class labels +, -. Build a two-level decision tree.

A	B	C	Number of Instances	
			+	−
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) (3 points) According to the classification error rate, which attribute would be chosen as the first splitting attribute? For each attribute, show the contingency table and the gains in classification error rate.

Classification error before split:

$$\text{CE} = 1 - \underset{i}{max}[p(i|t)] = 1 - \max\left[\frac{50}{100}, \frac{50}{100}\right] = \frac{50}{100}$$

Attribute A:

A	+	-
T	25	0
F	25	50

Contingency Table for A

$$CE_{A_T} = 1 - \max\left[\frac{25}{25}, \frac{0}{25}\right] = 0$$

$$CE_{A_F} = 1 - \max\left[\frac{25}{75}, \frac{50}{75}\right] = \frac{25}{75}$$

$$\Delta A = \frac{50}{100} - \left(0 \cdot \frac{25}{100} + \frac{25}{75} \cdot \frac{75}{100}\right) = \mathbf{\frac{75}{100}}$$

Attribute B:

B	+	-
T	30	20
F	20	30

Contingency Table for B

$$CE_{B_T} = 1 - \max\left[\frac{30}{50}, \frac{20}{50}\right] = \frac{20}{50}$$

$$CE_{B_F} = 1 - \max\left[\frac{20}{50}, \frac{30}{50}\right] = \frac{20}{50}$$

$$\Delta B = \frac{50}{100} - \left(\frac{20}{50} \cdot \frac{50}{100} + \frac{20}{50} \cdot \frac{50}{100}\right) = \mathbf{\frac{10}{100}}$$

Attribute C:

C	+	-
T	25	25
F	25	25

Contingency Table for C

$$CE_{C_T} = 1 - \max\left[\frac{25}{50}, \frac{25}{50}\right] = \frac{25}{50}$$

$$CE_{C_F} = 1 - \max\left[\frac{25}{50}, \frac{25}{50}\right] = \frac{25}{50}$$

$$\Delta C = \frac{50}{100} - \left(\frac{25}{50} \cdot \frac{50}{100} + \frac{25}{50} \cdot \frac{50}{100}\right) = \mathbf{0}$$

Since attribute A's gain is the highest ($\Delta A = 75/100$), would be chosen as the first split.

b) (3 points) Repeat for the two children of the root node. All instances of $A = T$ have a class label of '+', so we do not continue splitting.

We now look at entries for the $A = F$ child node:

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

Table 1: $A = F$ Child Node

We start with the classification error before further splitting:

$$EC = 1 - \max\left[\frac{25}{75}, \frac{50}{75}\right] = \frac{25}{75}$$

B	+	-
T	25	20
F	0	30

Contingency Table for B

$$CE_{B_T} = 1 - \max\left[\frac{25}{45}, \frac{20}{45}\right] = \frac{20}{45}$$

$$CE_{B_F} = 1 - \max\left[\frac{0}{30}, \frac{30}{30}\right] = 0$$

$$\Delta B = \frac{25}{75} - \left(\frac{20}{45} \cdot \frac{45}{75} + 0 \cdot \frac{30}{75}\right) = \mathbf{\frac{5}{75}}$$

$$CE_{C_T} = 1 - \max\left[\frac{0}{25}, \frac{25}{25}\right] = 0$$

C	+	-
T	0	25
F	25	25

Contingency Table for C

$$CE_{C_F} = 1 - \max\left[\frac{25}{50}, \frac{25}{50}\right] = \frac{25}{50}$$

$$\Delta C = \frac{25}{75} - \left(0 \cdot \frac{25}{75} + \frac{25}{50} \cdot \frac{50}{75}\right) = 0$$

Since $\Delta B > \Delta C$, B is chosen for the split.

c) (3 points) How many instances are misclassified by the resulting decision tree?

We would not need a node for $B = F$ so we just look at the $B = T$ node, which would have 20 misclassified instances.

d) (3 points) Repeat parts (a), (b), and (c) using C as the splitting attribute.

I don't think part (a) can be repeated if we already know we are splitting on C (since part (a) just asked us to determine which attribute to split on)? I'm assuming I should just follow the directions for parts (b) and (c).

Starting with the $C = T$ child node:

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

$C = T$ Child Node

$$EC = 1 - \max\left[\frac{25}{50}, \frac{25}{50}\right] = \frac{25}{50}$$

Attribute A:

A	+	-
T	25	0
F	0	25

Contingency Table for A

$$CE_{A_T} = 1 - \max\left[\frac{25}{25}, \frac{0}{25}\right] = 0$$

$$CE_{A_F} = 1 - \max\left[\frac{0}{25}, \frac{25}{25}\right] = 0$$

$$\Delta A = \frac{25}{50} - \left(0 \cdot \frac{25}{50} + 0 \cdot \frac{25}{50}\right) = \mathbf{\frac{25}{50}}$$

Attribute B:

B	+	-
T	5	20
F	20	5

Contingency Table for B

$$CE_{B_T} = 1 - \max\left[\frac{5}{25}, \frac{20}{25}\right] = \frac{5}{25}$$

$$CE_{B_F} = 1 - \max\left[\frac{20}{25}, \frac{5}{25}\right] = \frac{5}{25}$$

$$\Delta B = \frac{25}{50} - \left(\frac{5}{25} \cdot \frac{25}{50} + \frac{5}{25} \cdot \frac{25}{50}\right) = \mathbf{\frac{15}{50}}$$

We would choose **A** for the next split.

C=F child node: $EC = \frac{25}{50}$

Attribute A:

A	+	-
T	0	0
F	25	25

Contingency Table for A

$$CE_{A_T} = 0$$

$$CE_{A_F} = \frac{25}{50}$$

$$\Delta A = \frac{25}{50} - \left(0 \cdot \frac{0}{50} + \frac{25}{50} \cdot \frac{50}{50}\right) = \mathbf{0}$$

Attribute B:

B	+	-
T	25	0
F	0	25

Contingency Table for B

$$CE_{B_T} = 0$$
$$CE_{B_F} = 0$$
$$\Delta A = \mathbf{\frac{25}{50}}$$

Here we would split on B.

There are no misclassified instances resulting from this tree.

e) (3 points) Use the results in parts (c) and (d) to conclude about the greedy nature of the decision tree induction algorithm.

The difference we see between a split starting with A and a split starting with C illustrates the greedy nature of this algorithm. Splitting with A was a local solution, meaning that it reduced the classification error more than splitting with other features. It did not, however, result in the global solution found when starting to split with C.

Exercise 3.12

10 points

Consider a labeled data set containing 100 data instances, which is randomly partitioned into two sets A and B, each containing 50 instances. We use A as the training set to learn two decision trees T10 with 10 leaf nodes and T100 with 100 leaf nodes. The accuracies of the two decision trees on data sets A and B are shown below:

Data Set	T_{10}	T_{100}
A	0.86	0.97
B	0.84	0.77

a) (5 points) Based on the accuracies shown in the table above, which classification model would you expect to have better performance on unseen instances?

I would expect T_{10} to perform better on unseen instances. That is what we see in the chart above, where T_{10} had lower training accuracy than T_{100}

but higher testing accuracy. T_{100} appears to have overfit to its training data, which is not surprising given the higher number of trees.

b) (5 points) Now, you tested T_{10} and T_{100} on the entire data set $(A+B)$ and found that the classification accuracy of T_{10} on data set $(A+B)$ is 0.85, whereas the classification accuracy of T_{100} on the data set $(A+B)$ is 0.87. Based on this new information and your observations from Table 3.7, which classification model would you finally choose for classification?

   This wouldn't impact my decision to pick T_{10}. Since both models were trained on A, the results of testing on A are not meaningful and I would just base my decision on the results of testing on B alone.

Exercise 3.13

10 points

Consider the following approach for testing whether a classifier A beats another classifier B. Let N be the size of a given dataset, p_A be the accuracy of classifier A, p_B be the accuracy of classifier B, and $p = (p_A + p_B)/2$ be the average accuracy for both classifiers. To test whether classifier A is significantly better than B, the following Z-statistic is used:

$$Z = \frac{p_A - p_B}{\sqrt{\frac{2p(1-p)}{N}}}$$

Classifier A is assumed to be better than classifier B if $Z > 1.96$. Table 3.8 compares the accuracies of three different classifiers, decision tree classifiers, naive Bayes classifiers, and support vector machines, on various datasets. (The latter two classifiers are described in Chapter 4.)

Summarize the performance of the classifiers given in Table 3.8.

To answer this question I wrote a quick function to implement the equation above to calculate z scores for each combination of models/datasets. The function then compared z scores to the 1.96 threshold and outputs which classifier 'beat' the other classifier or if the two were equal ($-1.96 < z < 1.96$). I included a picture of the function I used at the end of the document just in case.

We can observe that the SVM classifier was the most likely to be equal or better to the other models. The Bayes classifier was the most likely to be worse than another model. A summary of model comparisons is shown in Table 2.

Exercise 3.14

7 points

Let X be a binomial random variable with mean, Np, and variance, $Np(1-p)$. Show that the ratio X/N also has a binomial distribution with mean, p, and variance, $p(1-p)/N$

Data Set	Size (N)	Decision Tree (%)	naïve Bayes (%)	Support vector machine (%)
Anneal	898	92.09	79.62	87.19
Australia	690	85.51	76.81	84.78
Auto	205	81.95	58.05	70.73
Breast	699	95.14	95.99	96.42
Cleve	303	76.24	83.50	84.49
Credit	690	85.80	77.54	85.07
Diabetes	768	72.40	75.91	76.82
German	1000	70.90	74.70	74.40
Glass	214	67.29	48.59	59.81
Heart	270	80.00	84.07	83.70
Hepatitis	155	81.94	83.23	87.10
Horse	368	85.33	78.80	82.61
Ionosphere	351	89.17	82.34	88.89
Iris	150	94.67	95.33	96.00
Labor	57	78.95	94.74	92.98
Led7	3200	73.34	73.16	73.56
Lymphography	148	77.03	83.11	86.49
Pima	768	74.35	76.04	76.95
Sonar	208	78.85	69.71	76.92
Tic-tac-toe	958	83.72	70.04	98.33
Vehicle	846	71.04	45.04	74.94
Wine	178	94.38	96.63	98.88
Zoo	101	93.07	93.07	96.04

Table 3.8

$$\mu_{\frac{X}{N}} = E\left[\frac{X}{N}\right] = \frac{1}{N}E[X] = \frac{1}{N}(Np) = p$$

If we define variance as $Var(X) = E[(X - E[X]^2] = Np(1-p)$, and we know from above that $E\left[\frac{X}{N}\right] = p$ we know that:

$$\begin{aligned}
\sigma^2_{\frac{X}{N}} = Var\left(\frac{X}{N}\right) &= E\left[\left(\frac{X}{N} - E\left[\frac{X}{N}\right]\right)^2\right] && (4)\\
&= E[(X - E[X]^2]/N^2 && (5)\\
&= Np(1-p)/N^2 && (6)\\
&= p(1-p)/N && (7)
\end{aligned}$$

Additional Material

		Decision Tree	Bayes	SVM	Total
Decision Tree	Better	0	10	2	12
	Equal	0	11	15	26
	Worse	0	2	6	8
Bayes	Better	2	0	0	2
	Equal	11	0	15	26
	Worse	10	0	8	18
SVM	Better	6	8	0	14
	Equal	15	15	0	30
	Worse	2	0	0	2

Table 2: Comparison of Classifiers

```python
def calculate_and_compare_z_score(data, classifier_1, classifier_2):
    p_a = data[classifier_1] / 100
    p_b = data[classifier_2] / 100
    n = data['Size']

    comparison = []
    for i in range(len(data)):
        if p_a[i] == p_b[i]:
            comparison.append('Equal')
        else:
            p = (p_a[i] + p_b[i]) / 2
            standard_error = np.sqrt((p * (1 - p)) / n[i])
            if standard_error == 0:
                comparison.append('Equal')
            else:
                z = (p_a[i] - p_b[i]) / (standard_error * np.sqrt(2))
                if z > 1.96:
                    comparison.append(classifier_1)
                elif z < -1.96:
                    comparison.append(classifier_2)
                else:
                    comparison.append('Equal')
    data[f'{classifier_1} vs {classifier_2}'] = comparison
    return data

data = calculate_and_compare_z_score(data, 'Decision Tree', 'Bayes')
data = calculate_and_compare_z_score(data, 'Decision Tree', 'SVM')
data = calculate_and_compare_z_score(data, 'Bayes', 'SVM')
```

The Z score function used to answer question 3.13