CS446: Machine Learning	**Fall 2014**

Problem Set 2

Chaitanya Datye *Handed In: September 26, 2014*

1. `Problem 1 Solution`

   (a)

   Let S be a collection of the 16 examples out of which we have $7+$ and $9-$ examples.

   We thus have $S \leftarrow [7+, 9-]$
   $Entropy(S) = -\frac{7}{16}\log_2(\frac{7}{16}) - \frac{9}{16}\log_2(\frac{9}{16}) = 0.988$

   Now, lets calculate the gain for each of the attributes.
   Color: Suppose S_Y and S_P are the collections of examples with color as Yellow and Purple respectively, we have
   $S_Y \leftarrow [5+, 3-]$ and $S_P \leftarrow [2+, 6-]$

   $$Entropy(S_Y) = -\frac{5}{8}\log_2(\frac{5}{8}) - \frac{3}{8}\log_2(\frac{3}{8}) = 0.9544$$
   $$Entropy(S_P) = -\frac{2}{8}\log_2(\frac{2}{8}) - \frac{6}{8}\log_2(\frac{6}{8}) = 0.8113$$
   $$Gain(S, Color) = E(S) - \frac{1}{2} \times 0.9544 - \frac{1}{2} \times 0.8113 = 0.1051$$

   Size: Suppose S_S and S_L are the collections of examples with size as Small and Large respectively, we have
   $S_S \leftarrow [5+, 3-]$ and $S_L \leftarrow [2+, 6-]$

   $$Entropy(S_S) = -\frac{5}{8}\log_2(\frac{5}{8}) - \frac{3}{8}\log_2(\frac{3}{8}) = 0.9544$$
   $$Entropy(S_L) = -\frac{2}{8}\log_2(\frac{2}{8}) - \frac{6}{8}\log_2(\frac{6}{8}) = 0.8113$$
   $$Gain(S, Size) = E(S) - \frac{1}{2} \times 0.9544 - \frac{1}{2} \times 0.8113 = 0.1051$$

   Act: Suppose S_S and S_D are the collections of examples with act as Stretch and Dip respectively, we have
   $S_S \leftarrow [5+, 3-]$ and $S_D \leftarrow [2+, 6-]$

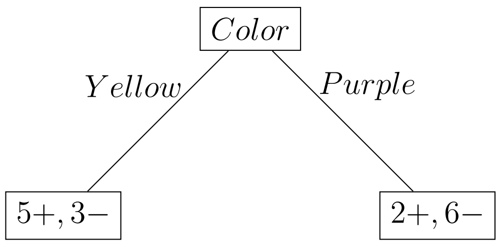
   $$Entropy(S_S) = -\frac{5}{8}\log_2(\frac{5}{8}) - \frac{3}{8}\log_2(\frac{3}{8}) = 0.9544$$
   $$Entropy(S_D) = -\frac{2}{8}\log_2(\frac{2}{8}) - \frac{6}{8}\log_2(\frac{6}{8}) = 0.8113$$
   $$Gain(S, Act) = E(S) - \frac{1}{2} \times 0.9544 - \frac{1}{2} \times 0.8113 = 0.1051$$

   Age: Suppose S_A and S_C are the collections of examples with age as Adult and Child respectively, we have
   $S_A \leftarrow [5+, 3-]$ and $S_C \leftarrow [2+, 6-]$

$$
\begin{aligned}
Entropy(S_A) &= -\tfrac{5}{8}\log_2(\tfrac{5}{8}) - \tfrac{3}{8}\log_2(\tfrac{3}{8}) = 0.9544 \\
Entropy(S_C) &= -\tfrac{2}{8}\log_2(\tfrac{2}{8}) - \tfrac{6}{8}\log_2(\tfrac{6}{8}) = 0.8113 \\
Gain(S, Age) &= E(S) \ - \ \tfrac{1}{2} \times 0.9544 \ - \ \tfrac{1}{2} \times 0.8113 \ = \ 0.1051
\end{aligned}
$$

Since we have the same Gain value for all the attributes, we break ties arbitrarily. Lets choose ***Color*** as our root node.

Let $Color = Yellow$ be the left branch and $Color = Purple$ be the right branch of the node. First level of the tree is done. Lets move on to the second level.



Now, at the left node where $Color = Yellow$, lets consider S_Y as a collection of all examples where $Color = Yellow$.

$$
\begin{aligned}
&S_Y \ \leftarrow [5+, 3-] \\
&Entropy(S_Y) \ = \ -\tfrac{5}{8}\log_2(\tfrac{5}{8}) - \tfrac{3}{8}\log_2(\tfrac{3}{8}) = 0.9544
\end{aligned}
$$

Now, lets calculate the gain for each of the remaining attributes.

Size: Suppose S_{SY} and S_{LY} are the collections of examples with $Color = Yellow$ and size as Small and Large respectively, we have
$S_{SY} \ \leftarrow [4+, 0-]$ and $S_{LY} \ \leftarrow [1+, 3-]$

$$
\begin{aligned}
Entropy(S_{SY}) &= 0 \\
Entropy(S_{LY}) &= -\tfrac{1}{4}\log_2(\tfrac{1}{4}) - \tfrac{3}{4}\log_2(\tfrac{3}{4}) = 0.8113 \\
Gain(S_Y, Size) &= E(S_Y) \ - \ \tfrac{1}{2} \times 0 \ - \ \tfrac{1}{2} \times 0.8113 \ = \ 0.5488
\end{aligned}
$$

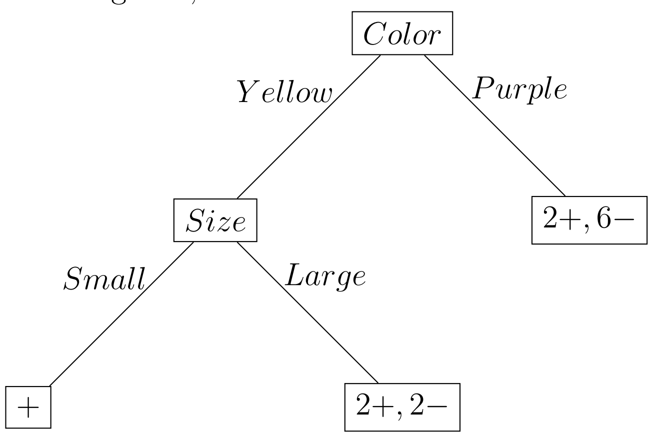
Act: Suppose S_{SY} and S_{DY} are the collections of examples with $Color = Yellow$ and act as Stretch and Dip respectively, we have
$S_{SY} \ \leftarrow [3+, 1-]$ and $S_{DY} \ \leftarrow [2+, 2-]$

$$
\begin{aligned}
Entropy(S_{SY}) &= -\tfrac{3}{4}\log_2(\tfrac{3}{4}) - \tfrac{1}{4}\log_2(\tfrac{1}{4}) = 0.31128 \\
Entropy(S_{DY}) &= -\tfrac{2}{4}\log_2(\tfrac{2}{4}) - \tfrac{2}{4}\log_2(\tfrac{2}{4}) = 1 \\
Gain(S_Y, Act) &= E(S_Y) \ - \ \tfrac{1}{2} \times 0.31128 \ - \ \tfrac{1}{2} \times 1 \ = \ 0.2988
\end{aligned}
$$

Age: Suppose S_{AY} and S_{CY} are the collections of examples with $Color = Yellow$ and age as Adult and Child respectively, we have
$S_{AY} \ \leftarrow [3+, 1-]$ and $S_{CY} \ \leftarrow [2+, 2-]$

$$
\begin{aligned}
Entropy(S_{AY}) &= -\tfrac{3}{4}\log_2(\tfrac{3}{4}) - \tfrac{1}{4}\log_2(\tfrac{1}{4}) = 0.31128 \\
Entropy(S_{CY}) &= 1 \\
Gain(S_Y, Age) &= E(S_Y) - \tfrac{1}{2} \times 0.31128 - \tfrac{1}{2} \times 1 = 0.2988
\end{aligned}
$$

Since we can see that the value for $Gain(S_Y, Size)$ is the maximum among the three, we choose the attribute Size for the split in this level. We thus have the following tree,



Now, lets see what happens if the $Color = Yellow\ and\ Size = Large$. We have two attributes left - age and act.
Let S_{LY} denote the set of all examples with $Color = Yellow$ and $Size = Large$
$S_{LY} \leftarrow [1+, 3-]$
$Entropy(S_{LY}) = -\frac{1}{4}\log_2(\frac{1}{4}) - \frac{3}{4}\log_2(\frac{3}{4}) = 0.8113$
Now, lets calculate the gain for each of the remaining attributes.

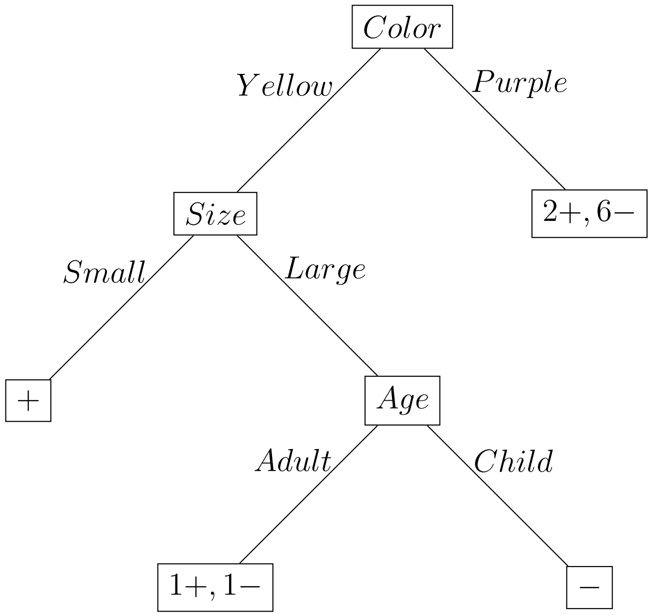
Act: Suppose S_{SLY} and S_{DLY} are the collections of examples with $Color = Yellow$, $Size = Large$ and act as Stretch and Dip respectively, we have
$S_{SLY} \leftarrow [1+, 1-]$ and $S_{DLY} \leftarrow [0+, 2-]$

$$
\begin{aligned}
Entropy(S_{SLY}) &= 1 \\
Entropy(S_{DLY}) &= 0 \\
Gain(S_{LY}, Act) &= E(S_{LY}) - \tfrac{1}{2} \times 1 - \tfrac{1}{2} \times 0 = 0.3113
\end{aligned}
$$

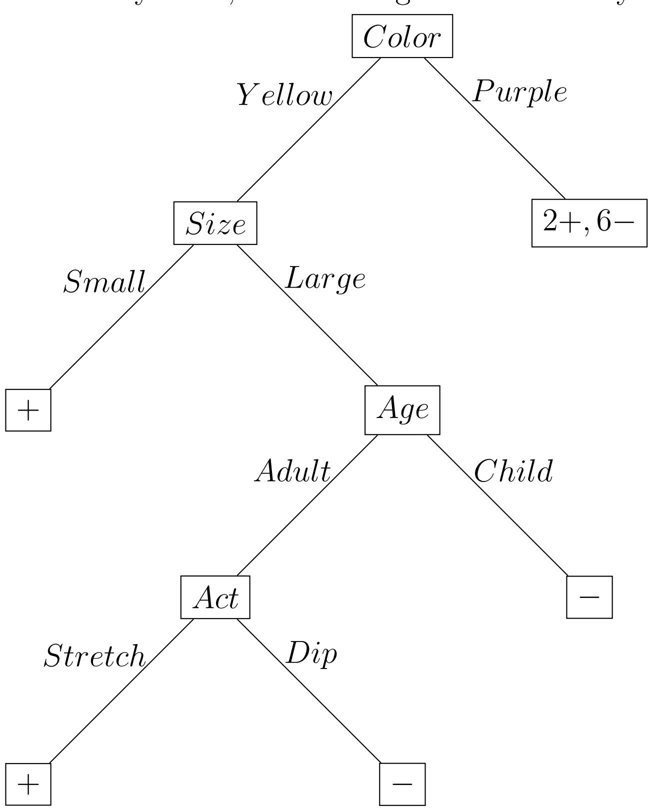
Age: Suppose S_{ALY} and S_{CLY} are the collections of examples with $Color = Yellow$, $Size = Large$ and age as Adult and Child respectively, we have
$S_{ALY} \leftarrow [1+, 1-]$ and $S_{CLY} \leftarrow [0+, 2-]$

$$
\begin{aligned}
Entropy(S_{ALY}) &= 1 \\
Entropy(S_{CLY}) &= 0 \\
Gain(S_{LY}, Age) &= E(S_{LY}) - \tfrac{1}{2} \times 1 - \tfrac{1}{2} \times 0 = 0.3113
\end{aligned}
$$

Since the *Gain* value for Act and Age are same, we break ties arbitrarily. Lets say we choose Age to make a decision at this step. Our tree then looks like,



Now, we just have one attribute left namely, Act. We can see that when $Color = Yellow,\ Size = Large\ and\ Age = Adult$, if $Act = Stretch$ then we have positive labels only. Else, we have negative labels only. We thus get the following tree.



We now consider what happens when $Color = Purple$. We have, $S_P \leftarrow [2+, 6-]$
$Entropy(S_P) \ = \ -\frac{2}{8}\log_2(\frac{2}{8}) - \frac{6}{8}\log_2(\frac{6}{8}) = 0.8113$

Lets calculate the Gain for the other attributes,

Size: Suppose S_{SP} and S_{LP} are the collections of examples with $Color = Purple$ and size as Small and Large respectively, we have
$S_{SP} \leftarrow [1+, 3-]$ and $S_{LP} \leftarrow [1+, 3-]$

$$\begin{aligned}
Entropy(S_{SP}) &= -\tfrac{1}{4}\log_2(\tfrac{1}{4}) - \tfrac{3}{4}\log_2(\tfrac{3}{4}) = 0.8113 \\
Entropy(S_{LP}) &= -\tfrac{1}{4}\log_2(\tfrac{1}{4}) - \tfrac{3}{4}\log_2(\tfrac{3}{4}) = 0.8113 \\
Gain(S_P, Size) &= E(S_P) - \tfrac{1}{2} \times 0.8113 - \tfrac{1}{2} \times 0.8113 = 0
\end{aligned}$$

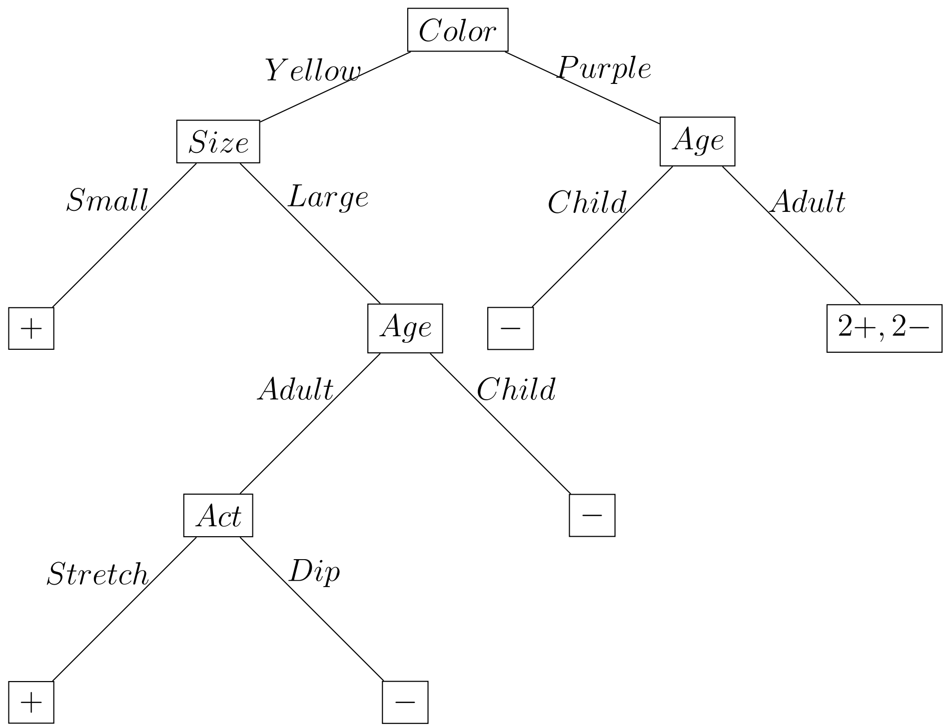
Act: Suppose S_{SP} and S_{DP} are the collections of examples with $Color = Purple$ and Act as Stretch and Dip respectively, we have
$S_{SP} \leftarrow [2+, 2-]$ and $S_{DP} \leftarrow [0+, 4-]$

$$\begin{aligned}
Entropy(S_{SP}) &= 1 \\
Entropy(S_{DP}) &= 0 \\
Gain(S_P, Act) &= E(S_P) - \tfrac{1}{2} \times 1 - \tfrac{1}{2} \times 0 = 0.3113
\end{aligned}$$

Age: Suppose S_{AP} and S_{CP} are the collections of examples with $Color = Purple$ and age as Adult and Child respectively, we have
$S_{AP} \leftarrow [2+, 2-]$ and $S_{CP} \leftarrow [0+, 4-]$

$$\begin{aligned}
Entropy(S_{AP}) &= 1 \\
Entropy(S_{CP}) &= 0 \\
Gain(S_P, Age) &= E(S_P) - \tfrac{1}{2} \times 1 - \tfrac{1}{2} \times 0 = 0.3113
\end{aligned}$$

Since we have a tie, we break it arbitrarily. Lets choose Age for making a decision. We thus have the following tree.



We now consider the remaining two attributes namely Size and Act. Let S_{AP} denote a set where $Color = Purple$ and $Age = Adult$. We have, $S_{AP} \leftarrow [2+, 2-]$
$Entropy(S_{AP}) = 1$

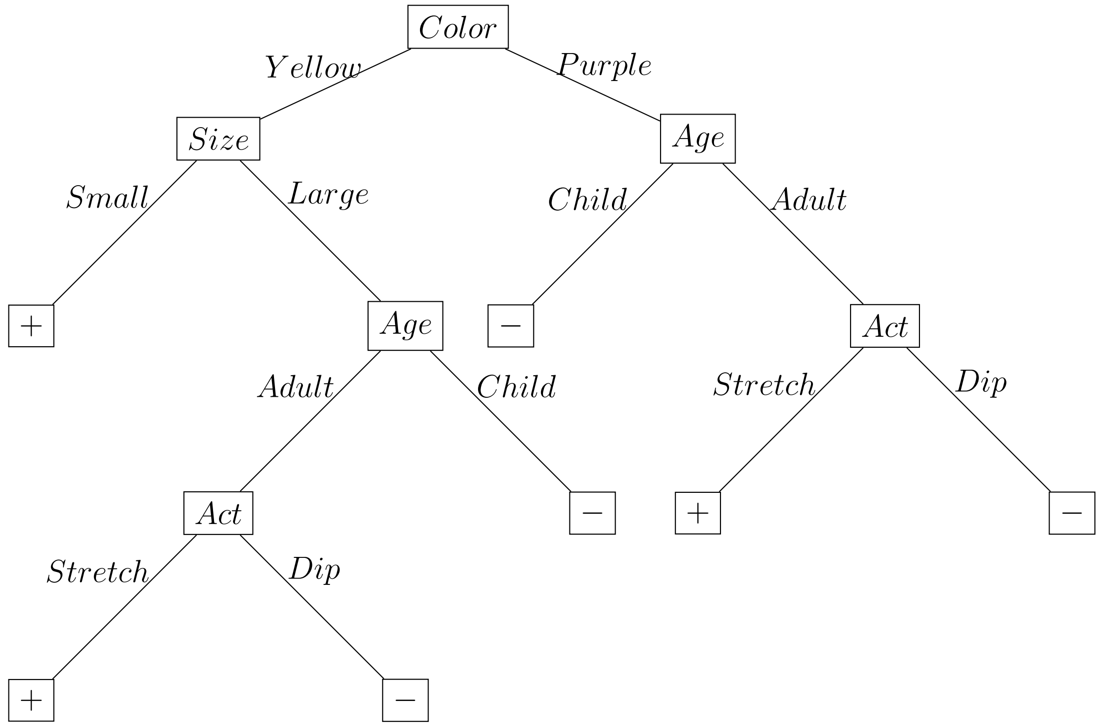
Size: Suppose S_{SAP} and S_{LAP} are the collections of examples with $Color = Purple$, $Age = Adult$ and size as Small and Large respectively, we have
$S_{SAP} \leftarrow [1+, 1-]$ and $S_{LAP} \leftarrow [1+, 1-]$

$Entropy(S_{SAP}) = 1$
$Entropy(S_{LAP}) = 1$
$Gain(S_{AP}, Size) = E(S_{AP}) - \frac{1}{2} \times 1 - \frac{1}{2} \times 1 = 0$

Act: Suppose S_{SAP} and S_{DAP} are the collections of examples with $Color = Purple$, $Age = Adult$ and Act as Stretch and Dip respectively, we have
$S_{SAP} \leftarrow [2+, 2-]$ and $S_{DAP} \leftarrow [0+, 2-]$

$Entropy(S_{SAP}) = 0$
$Entropy(S_{DAP}) = 0$
$Gain(S_{AP}, Act) = E(S_{AP}) - \frac{1}{2} \times 0 - \frac{1}{2} \times 0 = 1$

We thus choose Act as our decision variable here. We get the following tree//



(b)

Here we have everything similar to Part(a) except that the entropy function changes. Instead of Entropy, we use the $MajorityError$ function,
$MajorityError \ = \ \min(p, 1-p)$

We can redefine Gain as,
$Gain(S, A) \ = \ M(S) \ - \sum\limits_{v \in Values(A)} \frac{|S_v|}{|S|} \times M(S_v)$

Let S be a collection of the 16 examples out of which we have 7+ and 9− examples.

We thus have $S \leftarrow [7+, 9-]$
We have the initial error as,
$M(S) \ = \ min(\frac{7}{16}, \frac{9}{16}) \ = \ \frac{7}{16}$

Now, lets calculate the gain for each of the attributes.
Color: Suppose S_Y and S_P are the collections of examples with color as Yellow and Purple respectively, we have
$S_Y \ \leftarrow [5+, 3-]$ and $S_P \ \leftarrow [2+, 6-]$

$M(S_Y) \ = \ \frac{3}{8}$
$M(S_P) \ = \ \frac{2}{8}$
$Gain(S, Color) \ = \ M(S) \ - \ \frac{1}{2} \times \frac{3}{8} \ - \ \frac{1}{2} \times \frac{2}{8} \ = \ 0.125$

Size: Suppose S_S and S_L are the collections of examples with size as Small and

Large respectively, we have
$S_S \leftarrow [5+, 3-]$ and $S_L \leftarrow [2+, 6-]$

$M(S_S) = \frac{3}{8}$
$M(S_L) = \frac{2}{8}$
$Gain(S, Size) = M(S) - \frac{1}{2} \times \frac{3}{8} - \frac{1}{2} \times \frac{2}{8} = 0.125$

Act: Suppose S_S and S_D are the collections of examples with Act as Stretch and Dip respectively, we have
$S_S \leftarrow [5+, 3-]$ and $S_D \leftarrow [2+, 6-]$

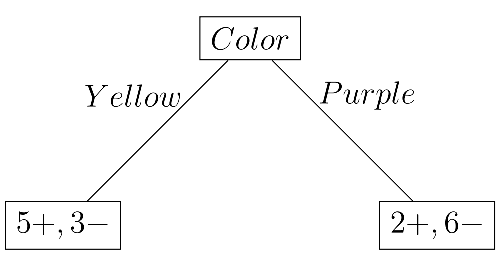
$M(S_S) = \frac{3}{8}$
$M(S_D) = \frac{2}{8}$
$Gain(S, Act) = M(S) - \frac{1}{2} \times \frac{3}{8} - \frac{1}{2} \times \frac{2}{8} = 0.125$

Age: Suppose S_A and S_C are the collections of examples with Age as Adult and Child respectively, we have
$S_A \leftarrow [5+, 3-]$ and $S_C \leftarrow [2+, 6-]$

$M(S_A) = \frac{3}{8}$
$M(S_C) = \frac{2}{8}$
$Gain(S, Age) = M(S) - \frac{1}{2} \times \frac{3}{8} - \frac{1}{2} \times \frac{2}{8} = 0.125$

Since we have the same Gain value for all the attributes, we break ties arbitrarily. Lets choose ***Color*** as our root node.

Let $Color = Yellow$ be the left branch and $Color = Purple$ be the right branch of the node. First level of the tree is done. Lets move on to the second level.



We consider Color and go for the gain of other attributes. We get exactly same steps as the decision tree in Part(a)

We proceed further in a similar fashion as Part(a) every time calculating the MajorityError function, we get the following Gain values at each step.

$M(S_Y) = \frac{3}{8}$
$Gain(S_Y, Size) = M(S_Y) - \frac{1}{2} \times 0 - \frac{1}{2} \times \frac{1}{4} = 0.125$
$Gain(S_Y, Act) = M(S_Y) - \frac{1}{2} \times \frac{1}{4} - \frac{1}{2} \times \frac{1}{2} = 0$

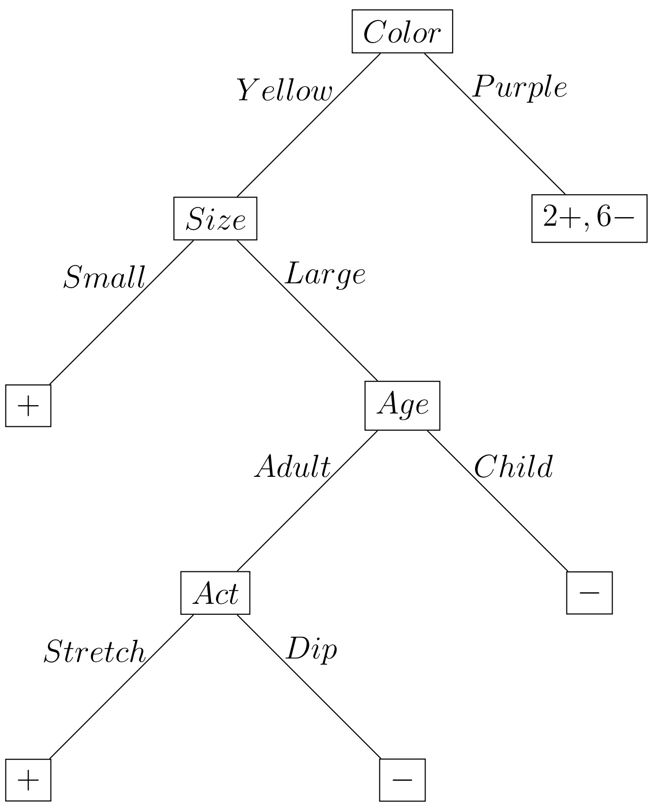
$Gain(S_Y, Age) \;=\; M(S_Y) \;-\; \frac{1}{2} \times \frac{1}{4} \;-\; \frac{1}{2} \times \frac{1}{2} \;=\; 0$

We choose Size as our deciding factor at this step. If $Size = Small$, we can set the class label to $+$, Else we compute further.

$M(S_{LY}) \;=\; \frac{1}{4}$
$Gain(S_{LY}, Act) \;=\; M(S_Y) \;-\; \frac{1}{2} \times \frac{1}{2} \;-\; \frac{1}{2} \times 0 \;=\; 0$
$Gain(S_{LY}, Age) \;=\; M(S_Y) \;-\; \frac{1}{2} \times \frac{1}{2} \;-\; \frac{1}{2} \times 0 \;=\; 0$

Since we have a tie here, we choose Age arbitrarily to make a decision. If $Color = Yellow$, $Size = Large$, $Age = Child$, we can classify this as a $-$. If $Age = Adult$, we check for Act. If $Act = Stretch \rightarrow +$. If $Act = Dip \rightarrow -$.

At this step, we get the following tree



Similarly, we go on the brach where $Color = Purple$ and compute Gain at each step, we get the following tree at the final step.

$M(S_P) \;=\; \frac{2}{8}$
$Gain(S_P, Size) \;=\; M(S_P) \;-\; \frac{1}{2} \times \frac{1}{4} \;-\; \frac{1}{2} \times \frac{1}{4} \;=\; 0$
$Gain(S_P, Act) \;=\; M(S_P) \;-\; \frac{1}{2} \times \frac{1}{2} \;-\; \frac{1}{2} \times 0 \;=\; 0$
$Gain(S_P, Age) \;=\; M(S_P) \;-\; \frac{1}{2} \times \frac{1}{2} \;-\; \frac{1}{2} \times 0 \;=\; 0$

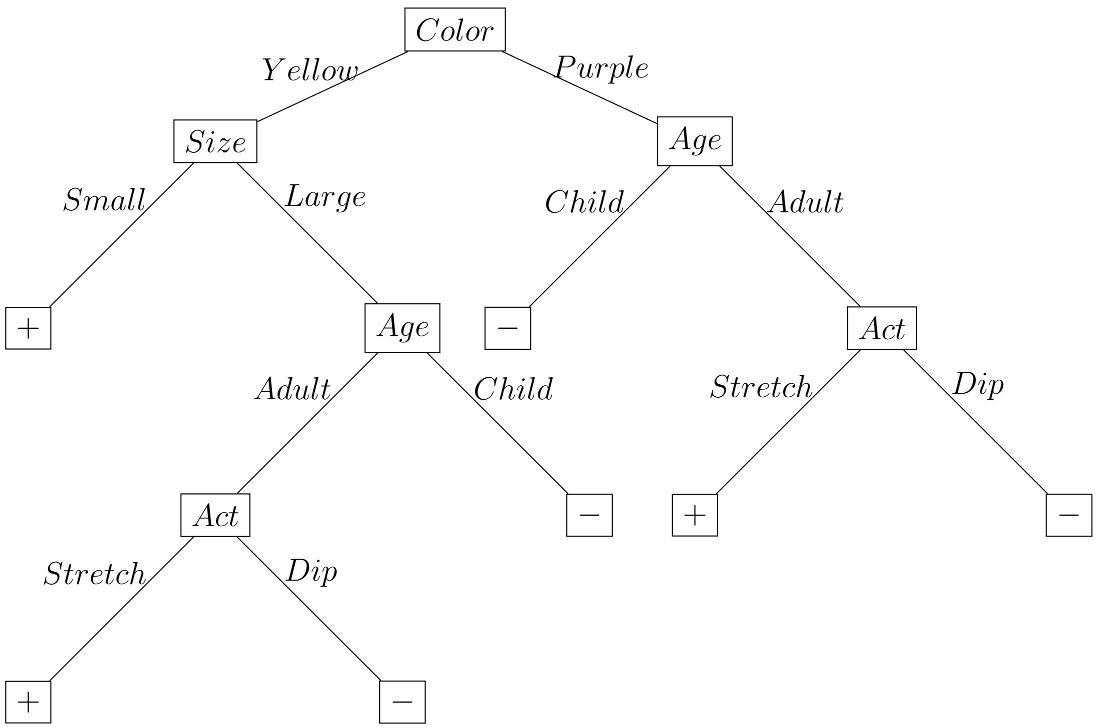
We break ties arbitrarily and choose Age as our decision function at this step.

$M(S_{AP}) \ = \ \frac{1}{2}$

$Gain(S_{AP}, Size) \ = \ M(S_{AP}) \ - \ \frac{1}{2} \times \frac{1}{2} \ - \ \frac{1}{2} \times \frac{1}{2} \ = \ 0$

$Gain(S_{AP}, Act) \ = \ M(S_{AP}) \ - \ \frac{1}{2} \times 0 \ - \ \frac{1}{2} \times 0 \ = \ 0.5$

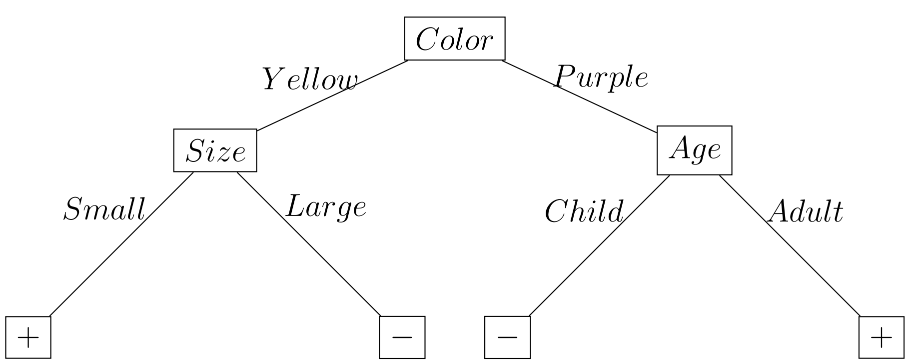
Hence, the final tree looks like,



`NOTE:` Since we have chosen to break ties similarly in $Part(a)$ and $Part(b)$, we are getting a similar tree. If we choose different attributes when we are breaking the ties when gains are same, we might get different trees.

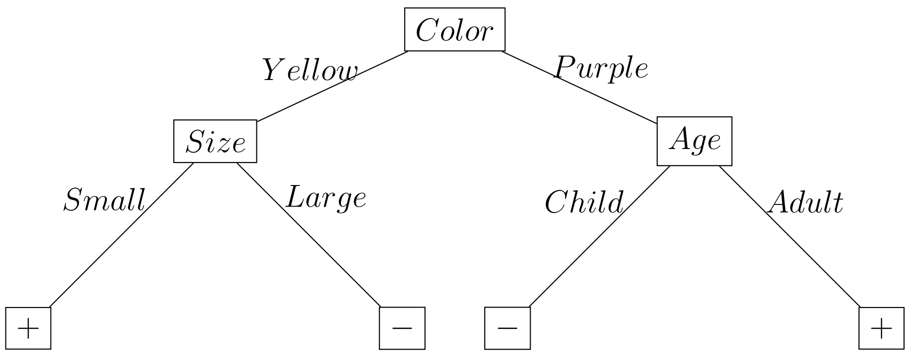
(c) We have to train our decision tree classifier on the first 12 examples and then evaluate it on the remaining 4 examples.

If we use Entropy function, we get the following tree for $Depth = 2$,



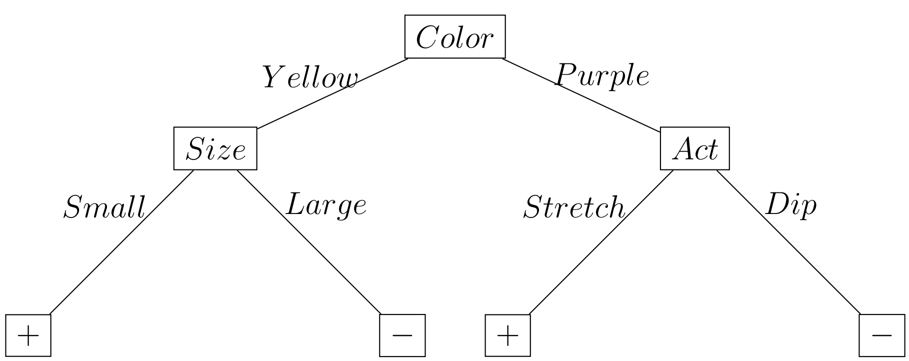
Using this tree, we evaluate the examples from example (13) *to* (16). We can see that, for example (14), $< Purple, Large, Dip, Adult >$, our tree would have predicted a $+$ value but its actual class label is $-$. So, this has been misclassified. Hence, the number of correctly classified examples is 3 $out\ of$ 4. Hence the error rate here is $E_r = \frac{1}{4} = 0.25$.

If we use Majority Error function, we get the following tree for $Depth = 2$. We break ties similarly as the previous part. We choose Color and Age when there is a tie. The tree looks like this,



Using this tree, we evaluate the examples from example (13) *to* (16). We can see that, for example (14), $< Purple, Large, Dip, Adult >$, our tree would have predicted a $+$ value but its actual class label is $-$. So, this has been misclassified. Hence, the number of correctly classified examples is 3 $out\ of$ 4. Hence the error rate here is $E_r = \frac{1}{4} = 0.25$.

`NOTE:` We have broken ties in a similar manner. If at the first stage we break ties using the Color attribute and in the second stage we break ties using the Act Attribute, instead of the Age attribute, we get the following tree.



Even in this case, if we use this tree to make predictions on our test data, we can see that the example 14 is classified wrongly. Hence, we still get the same error rate $= 0.25$.

2. `Problem 2 Solution`

(b) When we run the SGD algorithm on the training set and perform a 5-fold cross-validation, we get the following values of accuracy.
82%
81.75%
85.75%
80.25%
84.25%

Mean predictive accuracy is thus 82.8%

We thus have sample mean is $\bar{X} = 82.8$
Also, the sample standard deviation is $S = 2.1823$
Degrees of Freedom are $df = 4$
Number of examples are $n = 5$
From the t-table, for $df = 4$, we get $t_{\frac{\alpha}{2}} = t_{0.005} = 4.604$

We know that confidence interval is give by
$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

We can substitute and find that the confidence interval lies between
$\mathbf{99\%Confidence\ Interval} = (\mathbf{78.3067},\ \mathbf{87.2933})$

(c) When we train the ID3 classifier without limiting the depth on the training set and perform a 5-fold cross-validation, we get the following values of accuracy.
90%
88.5%
85.75%
87.75%
88.75%

Mean predictive accuracy is thus 88.15%

We thus have sample mean is $\bar{X} = 88.15$
Also, the sample standard deviation is $S = 1.5672$
Degrees of Freedom are $df = 4$
Number of examples are $n = 5$
From the t-table, for $df = 4$, we get $t_{\frac{\alpha}{2}} = t_{0.005} = 4.604$

We know that confidence interval is give by
$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

We can substitute and find that the confidence interval lies between
$\textbf{99\%Confidence Interval} = (\mathbf{84.923}, \ \mathbf{91.377})$

(d) i. ***Depth=4* -** When we train the ID3 classifier with maximum depth = 4 on the training set and perform a 5-fold cross-validation, we get the following values of accuracy.
77.75%
79%
76%
77.5%
77.5%

Mean predictive accuracy is thus 77.55%

We thus have sample mean is $\bar{X} = 77.55$
Also, the sample standard deviation is $S = 1.0665$
Degrees of Freedom are $df = 4$
Number of examples are $n = 5$
From the t-table, for $df = 4$, we get $t_{\frac{\alpha}{2}} = t_{0.005} = 4.604$

We know that confidence interval is give by
$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

We can substitute and find that the confidence interval lies between
$\textbf{99\%Confidence Interval} = (\mathbf{75.354}, \ \mathbf{79.746})$

ii. ***Depth=8* -** When we train the ID3 classifier with maximum depth = 8 on the training set and perform a 5-fold cross-validation, we get the following values of accuracy.
87%
81%
81%
83.75%
83.25%

Mean predictive accuracy is thus 83.2%

We thus have sample mean is $\bar{X} = 83.2$
Also, the sample standard deviation is $S = 2.4710$
Degrees of Freedom are $df = 4$
Number of examples are $n = 5$
From the t-table, for $df = 4$, we get $t_{\frac{\alpha}{2}} = t_{0.005} = 4.604$

We know that confidence interval is give by

$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

We can substitute and find that the confidence interval lies between
99%Confidence Interval $=$ (**78.1121**, **88.2878**)

(e) When we train the ID3 classifier without limiting the depth on the training set and perform a 5-fold cross-validation, we get the following values of accuracy.
86.75%
84.75%
86.75%
82.25%
84.75%

Mean predictive accuracy is thus 85.05%

We thus have sample mean is $\bar{X} = 85.05$
Also, the sample standard deviation is $S = 1.8574$
Degrees of Freedom are $df = 4$
Number of examples are $n = 5$
From the t-table, for $df = 4$, we get $t_{\frac{\alpha}{2}} = t_{0.005} = 4.604$

We know that confidence interval is give by
$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

We can substitute and find that the confidence interval lies between
99%Confidence Interval $=$ (**81.225**, **88.874**)

We thus get the following rankings on the classifiers based on the predictive accuracy value:

ID3 without specifying max_depth - performs better than SGD and Decision Stumps with $P_A = 88.15$. However, there might be a chance of overfitting. So, we cannot say whether this will be the best choice for the test data.
Decision Stumps - performs the Best on the given data with $P_A = 85.05$.
ID3 with max_depth(8) - performs better than SGD but not as good as Decision Stumps with $P_A = 83.2$
SGD - performs the less as compared to simple ID3 and ID3 with 8 as the max_depth and also as compared to Decision Stumps with with $P_A = 82.8$
ID3 with max_depth(4) - performs bad with respect to the other algorithms $P_A = 77.55$

We can see that the average accuracy of the ID3 run without specifying any maximum depth has the highest accuracy. But with pruned levels, the accuracy goes down a bit.

We can thus say that when an ID3 is constructed, even though the accuracy is high on training data, there might be a chance of overfitting of data and then ID3 algorithm may not perform that well on the test data. Hence, we cannot consider it as the best algorithm. As we put a restriction on how deep the tree can grow, we see a decrease in the accuracy. We see that the accuracy value when max_depth = 4 is even less than the value when max_depth = 8. So, the tree formed with ID3 with max_depth = 4, might not be perfect on the training data but might be able to classify the test data in a better fashion than the simple ID3 and the ID3 with max_depth = 8. But, with the decision stumps algorithm, we first generate 100 decision stumps of depth 4 and then use the stochastic gradient descent to calculate the accuracy. This seems to be a more accurate solution as it might perform very well on the test data. Also, since we are randomizing the training samples and then doing a sampling over the training data, it is better to use decision stumps as compared to decision tree. Since, using decision trees will prove to be computationally expensive. Since in case of stumps, the depth is fixed, the number of computations will be less as compared to the decision trees without any specified depth.

In case of decision stumps, even if some classifiers predict the data in a bad manner, these will be washed out in the final accuracy value as noise. Since, there will be many more classifiers predicting the values in a good way.

Hence, the Decision Stumps algorithm can be said to have performed the best on the dataset.