CIS 419/519: Homework 1

{Shruti Sinha}

Although the solutions are my own, I consulted with the following people while working on this homework: {Hemanth Vihari Kothapalli}

1. (a) Show your work:

$$P(\text{play outside} = \text{yes}) = p_+ = 20/45$$

$$P(\text{play outside} = \text{no}) = p_- = 25/45$$

We know,

$$Entropy = H(Y) = -p_+ log(p_+) - p_- log(p_-)$$

Therefore,

$$Entropy(playoutside) = -\frac{20}{45}log_2(\frac{20}{45}) - \frac{25}{45}log_2(\frac{25}{45})$$

$H(Y) = 0.9909$

$$IG = Entropy(S) - \sum_{v\epsilon values(S)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Lets denote Sunny=yes as X=1 and Sunny=no as X=0. We also denote play outside=yes with y=1, and play outside = no with y=0 Therefore for attribute: sunny,

$$P(X = 0)_{sunny} = \frac{24}{45},$$

$$P(X = 1)_{sunny} = \frac{21}{45}$$

For sunny,

$P(y = 1|X = 1) = \frac{15}{21}, P(y = 1|X = 0) = \frac{5}{24}, P(y = 0|X = 1)\frac{6}{21}, P(y = 0|X = 0) = \frac{19}{24}$

$$H(y|X) = -\frac{21}{45}E(S_{Sunny=yes}) - \frac{24}{45}E(S_{Sunny=no}) \qquad (1)$$

$$H(y|X) = -\frac{21}{45}(\frac{15}{21}log_2\frac{15}{21} + \frac{6}{21}log_2\frac{6}{21}) - \frac{24}{45}(\frac{5}{24}log_2\frac{5}{24} + \frac{19}{24}log_2\frac{19}{24}) \tag{2}$$

$$H(y|X) = 0.79656 \tag{3}$$

$IG_{sunny} = H(Y) - H(y|X)$
$IG_{sunny} = 0.9909 - 0.796565$
$IG_{sunny} = 0.1943332$

Now, for attribute as Snow.
Lets denote Snow=yes as X=1 and Snow=no asX=0. We also denote play outside=yes with y=1, and play outside=no with y=0 Therefore for attribute: snow,

$$P(X=0)_{snow} = \frac{37}{45},$$

$$P(X=1)_{snow} = \frac{8}{45}$$

For snow,

$$P(y=1|X=1) = \frac{6}{8}, P(y=1|X=0) = \frac{14}{37}, P(y=0|X=1)\frac{2}{8}, P(y=0|X=0) = \frac{23}{37}$$

$$H(y|X) = -\frac{37}{45}E(S_{Snow=no}) + -\frac{8}{45}E(S_{Snow=yes})$$

$$H(y|X) = -\frac{37}{45}(\frac{14}{37}log_2\frac{14}{37} + \frac{23}{37}log_2\frac{23}{37}) - \frac{8}{45}(\frac{2}{8}log_2\frac{2}{8} + \frac{6}{8}log_2\frac{6}{8})$$

$$H(y|X) = 0.931070845$$

$IG_{snow} = H(Y) - H(y|X)$
$IG_{snow} = 0.9909 - 0.931070845$
$IG_{snow} = 0.0598291$
Therefore,

$$IG_{Snow} = 0.0598291$$

$$IG_{Sunny} = 0.1943332$$

Sunny is better because we get its IG higher than that of Snow.Hence,we pick Sunny as root attribute.

(b) According to the given equation we consider,

$$Majority\,Error(Y) = min(p, 1-p)$$

where, p = fraction of examples labelled T = $\frac{8}{16}$
1-p = fraction of examples labelled F = $\frac{8}{16}$

$$Majority\,Error(Y) = min(p, 1-p) = \frac{8}{16}$$

$$IG = Majority\, Error(Y) - Majority\, Error(Y|X)$$

$$MajorityError(y|X) = P(X=0)MajorityError(y_{X=0}) + P(X=1)MajorityError(y_{X=1})$$

We consider first attribute as color.
Lets denote color= blue,X=0 and color = red,X=1.
$P(X=0)_{color} = \frac{9}{16}$, $P(X=1)_{color} = \frac{7}{16}$ We also denote play outside = yes with y=1, and play outside = no with y=0 For color,

$$P(y=1|X=1) = \frac{4}{7}, P(y=1|X=0) = \frac{4}{9}, P(y=0|X=1) = \frac{3}{7}, P(y=0|X=0) = \frac{5}{9}$$

$$MajorityError(y|X) = \frac{9}{16}MajorityError(y_{X=0}) + \frac{7}{16}(MajorityError(y_{X=1}))$$

$$MajorityError(y|X) = \frac{9}{16}min(\frac{4}{9}, \frac{5}{9}) + \frac{7}{16}min(\frac{4}{7}, \frac{3}{7})$$

$$MajorityError(y|X)_{color} = \frac{9}{16} * \frac{4}{9} + \frac{7}{16} * \frac{3}{7}$$

$$IG_{color} = \frac{8}{16} - \frac{7}{16} = \frac{1}{16}$$

We consider attribute as size.
Lets denote size=small,X=0 and size = large,X=1.
$P(X=0)_{small} = \frac{9}{16}$, $P(X=1)_{large} = \frac{7}{16}$ We also denote fraction labelled T with y=1, and fraction labelled F with y=0 For size,

$$P(y=1|X=1) = \frac{4}{7}, P(y=1|X=0) = \frac{4}{9}, P(y=0|X=1) = \frac{3}{7}, P(y=0|X=0) = \frac{5}{9}$$

$$MajorityError(y|X) = \frac{9}{16}MajorityError(y_{X=0}) + \frac{7}{16}(MajorityError(y_{X=1}))$$

$$MajorityError(y|X)_{size} = \frac{9}{16}min(\frac{4}{9}, \frac{5}{9}) + \frac{7}{16}min(\frac{4}{7}, \frac{3}{7})$$

$$MajorityError(y|X)_{size} = \frac{9}{16} * \frac{4}{9} + \frac{7}{16} * \frac{3}{7} = \frac{7}{16}$$

$$IG_{color} = \frac{8}{16} - \frac{7}{16} = \frac{1}{16}$$

We next consider attribute as Act.
Lets denote Act = Stretch,X=0 and Act=Dip,X=1.
$P(X=0)_{Stretch} = \frac{7}{16}$, $P(X=1)_{Dip} = \frac{9}{16}$ We also denote fraction labelled T with y=1, and fraction labelled F with y=0 For Act,

$$P(y=1|X=1) = \frac{4}{9}, P(y=1|X=0) = \frac{4}{7}, P(y=0|X=1) = \frac{5}{9}, P(y=0|X=0) = \frac{3}{7}$$

$$MajorityError(y|X) = \frac{7}{16}MajorityError(y_{X=0}) + \frac{9}{16}(MajorityError(y_{X=1}))$$

$$MajorityError(y|X)_{Act} = \frac{7}{16}min(\frac{4}{7}, \frac{3}{7}) + \frac{9}{16}min(\frac{4}{9}, \frac{5}{9})$$

$$MajorityError(y|X)_{Act} = \frac{7}{16} * \frac{3}{7} + \frac{9}{16} * \frac{4}{9} = \frac{7}{16}$$

$$IG_{Act} = \frac{8}{16} - \frac{7}{16} = \frac{1}{16}$$

We finally consider attribute as Age.
Lets denote Age = Adult,X=0 and Act=Child,X=1.
$P(X = 0)_{Adult} = \frac{8}{16}$, $P(X = 1)_{Child} = \frac{8}{16}$ We also denote fraction labelled T with y=1, and fraction labelled F with y=0 For Age,

$$P(y = 1|X = 1) = \frac{5}{8}, P(y = 1|X = 0) = \frac{3}{8}, P(y = 0|X = 1) = \frac{3}{8}, P(y = 0|X = 0) = \frac{5}{8}$$

$$MajorityError(y|X)_{age} = \frac{8}{16}MajorityError(y_{X=0}) + \frac{8}{16}(MajorityError(y_{X=1}))$$

$$MajorityError(y|X)_{Age} = \frac{8}{16}min(\frac{3}{8}, \frac{5}{8}) + \frac{8}{16}min(\frac{5}{8}, \frac{3}{8})$$

$$MajorityError(y|X)_{Age} = \frac{8}{16} * \frac{3}{8} + \frac{8}{16} * \frac{3}{8} = \frac{6}{16}$$

$$IG_{Act} = \frac{8}{16} - \frac{6}{16} = \frac{2}{16}$$

$$IG_{color} = 0.0625$$

$$IG_{size} = 0.0625$$

$$IG_{Act} = 0.0625$$

$$IG_{Age} = 0.125$$

We therefore pick Age as our root attribute and split the decision tree on it.
Moving on to greater depth we first consider Age = Adult and get leaf nodes for it.
We observe that Age = Adult has 3T and 5F.

Color	Size	Act	Age	Inflated	Count
Blue	Small	Stretch	Adult	F	2
Blue	Small	Dip	Adult	T	1
Blue	Large	Dip	Adult	F	1
Red	Small	Dip	Adult	F	2
Red	Large	Stretch	Adult	T	2

According to the above table, we first find out Majority Error(Y) for Age = Adult.

$$Majorityerror(Y) = min(p, 1 - p)$$

where, p = fraction of examples labelled T when Age = Adult = $\frac{3}{8}$
1-p = fraction of examples labelled F when Age = adult= $\frac{5}{8}$

$$MajorityError(Y) = min(p, 1 - p) = \frac{3}{8}$$

We consider attribute color first.
Lets denote color = blue,X=0 and color=red,X=1.
$P(X=0)_{blue} = \frac{4}{8}$, $P(X=1)_{red} = \frac{4}{8}$ We also denote fraction labelled T for Age= adult with y=1, and fraction labelled F for age = Adult with y=0 For Color,

$$P(y=1|X=1) = \frac{2}{4}, P(y=1|X=0) = \frac{1}{4}, P(y=0|X=1) = \frac{2}{4}, P(y=0|X=0) = \frac{3}{4}$$

$$MajorityError(y|X)_{color} = \frac{4}{8}MajorityError(y_{X=0}) + \frac{4}{8}(MajorityError(y_{X=1}))$$

$$MajorityError(y|X)_{color} = \frac{4}{8}min(\frac{3}{4}, \frac{1}{4}) + \frac{4}{8}min(\frac{2}{4}, \frac{2}{4})$$

$$MajorityError(y|X)_{color} = \frac{4}{8} * \frac{1}{4} + \frac{4}{8} * \frac{2}{4} = \frac{3}{8}$$

$$IG_{Color,Age=Adult} = \frac{3}{8} - \frac{3}{8} = 0$$

We consider attribute size next.
Lets denote size = small,X=0 and size=large,X=1.
$P(X=0)_{small} = \frac{5}{8}$, $P(X=1)_{large} = \frac{3}{8}$ We also denote fraction labelled T for Age= adult with y=1, and fraction labelled F for age = Adult with y=0 For Size,

$$P(y=1|X=1) = \frac{2}{3}, P(y=1|X=0) = \frac{1}{5}, P(y=0|X=1) = \frac{1}{3}, P(y=0|X=0) = \frac{4}{5}$$

$$MajorityError(y|X)_{size} = \frac{5}{8}min(\frac{1}{5}, \frac{4}{5}) + \frac{3}{8}min(\frac{2}{3}, \frac{1}{3})$$

$$MajorityError(y|X)_{size} = \frac{5}{8} * \frac{1}{5} + \frac{3}{8} * \frac{1}{3} = \frac{2}{8}$$

$$IG_{Size,Age=Adult} = \frac{3}{8} - \frac{2}{8} = \frac{1}{8}$$

We consider attribute Act next.
Lets denote Act = Stretch,X=0 and Act=Dip,X=1.
$P(X=0)_{Stretch} = \frac{4}{8}$, $P(X=1)_{Dip} = \frac{4}{8}$
For Act,

$$P(y=1|X=1) = \frac{1}{4}, P(y=1|X=0) = \frac{2}{4}, P(y=0|X=1) = \frac{3}{4}, P(y=0|X=0) = \frac{2}{4}$$

$$MajorityError(y|X)_{Act} = \frac{4}{8}min(\frac{2}{4}, \frac{2}{4}) + \frac{4}{8}min(\frac{1}{4}, \frac{3}{4})$$

$$MajorityError(y|X)_{Act} = \frac{4}{8} * \frac{2}{4} + \frac{4}{8} * \frac{1}{4} = \frac{3}{8}$$

$$IG_{Act,Age=Adult} = \frac{3}{8} - \frac{3}{8} = 0$$

$$IG_{color,Age=Adult} = 0$$

$$IG_{size,Age=Adult} = 0.125$$

$$IG_{Act,Age=Adult} = 0$$

Therefore, we split on size as the information gain is higher than the other two.
Referencing to the table above, we consider further try to find the leaf nodes for Age= adult and Size = Small.
We first find out Majority Error(Y) for Age = Adult and Size = Small.

$$Majorityerror(Y) = min(p, 1-p)$$

where, p = fraction of examples labelled T when Age = Adult,Size = Small = $\frac{1}{5}$
1-p = fraction of examples labelled F when Age = adult,Size = small = $\frac{4}{5}$

$$MajorityError(Y) = min(p, 1-p) = \frac{1}{5}$$

We consider attribute color first.
Lets denote color = blue,X=0 and color=red,X=1.
$P(X=0)_{blue} = \frac{3}{5}$, $P(X=1)_{red} = \frac{2}{5}$
For Color,

$$P(y=1|X=1) = \frac{0}{2}, P(y=1|X=0) = \frac{1}{3}, P(y=0|X=1) = \frac{2}{2}, P(y=0|X=0) = \frac{2}{3}$$

$$MajorityError(y|X)_{color} = \frac{3}{5}min(\frac{1}{3},\frac{2}{3}) + \frac{2}{5}min(\frac{2}{2},\frac{0}{2})$$

$$MajorityError(y|X)_{color} = \frac{1}{5}$$

$$IG_{Color,Age=Adult,Size=Small} = \frac{1}{5} - \frac{1}{5} = 0$$

We consider attribute Act next.
Lets denote Act = Stretch,X=0 and Act=Dip,X=1.
$P(X=0)_{Stretch} = \frac{2}{5}$, $P(X=1)_{Dip} = \frac{3}{5}$
For Act,

$$P(y=1|X=1) = \frac{1}{3}, P(y=1|X=0) = \frac{0}{2}, P(y=0|X=1) = \frac{2}{3}, P(y=0|X=0) = \frac{2}{2}$$

$$MajorityError(y|X)_{Act} = \frac{2}{5}min(\frac{0}{2},\frac{2}{2}) + \frac{3}{5}min(\frac{1}{3},\frac{2}{3})$$

$$MajorityError(y|X)_{Act} = \frac{1}{5}$$

$$IG_{Act,Age=Adult,Size=Small} = \frac{1}{5} - \frac{1}{5} = 0$$

Since there is a tie in IG for Act and Color, we pick color to split on since it appears first in the table. We split on color where, Color= Blue gives 2F and 1T Color= Red gives majority label as Inflated = F.
Therefore, for age = Adult, Size = Small and color = Blue we further look at Act.

It is seen that For color= blue, Size = Small and Age = Adult and Act = Dip we get Inflated = T and for Act = Stretch, Inflated = F. From the table we also conclude that there is tie between Color and Act for Age=Adult and Size = large.
We again pick Color as the attribute to split on since it appears first. Finally, for Age= Adult, Size= Large, we get that Color = Blue gives Inflated = F and Color = Red gives Inflated = T.

We now move onto Age = Child.The table for which is shown below.

Color	Size	Act	Age	Inflated	Count
Blue	Small	Dip	Child	T	3
Blue	Large	Stretch	Child	F	1
Blue	Large	Dip	Child	F	1
Red	Small	Dip	Child	F	1
Red	Large	Stretch	Child	T	2

According to the above table, we first find out Majority Error(Y) for Age = Child.

$$Majorityerror(Y) = min(p, 1-p)$$

where, p = fraction of examples labelled T when Age = Child = $\frac{5}{8}$
1-p = fraction of examples labelled F when Age = Child = $\frac{3}{8}$

$$MajorityError(Y) = min(p, 1-p) = \frac{3}{8}$$

We consider attribute color first.
Lets denote color = blue,X=0 and color=red,X=1.
$P(X=0)_{blue} = \frac{5}{8}$, $P(X=1)_{red} = \frac{3}{8}$ We also denote fraction labelled T for Age= Child with y=1, and fraction labelled F for Age = Child with y=0 For Color,

$$P(y=1|X=1) = \frac{2}{3}, P(y=1|X=0) = \frac{3}{5}, P(y=0|X=1) = \frac{1}{3}, P(y=0|X=0) = \frac{2}{5}$$

$$MajorityError(y|X)_{color} = \frac{5}{8}MajorityError(y_{X=0}) + \frac{3}{8}(MajorityError(y_{X=1}))$$

$$MajorityError(y|X)_{color} = \frac{5}{8}min(\frac{3}{5}, \frac{2}{5}) + \frac{3}{8}min(\frac{2}{3}, \frac{1}{3})$$

$$MajorityError(y|X)_{color} = \frac{5}{8} * \frac{2}{5} + \frac{3}{8} * \frac{1}{3} = \frac{3}{8}$$

$$IG_{Color,Age=Child} = \frac{3}{8} - \frac{3}{8} = 0$$

We consider attribute size next.
Lets denote size = small,X=0 and size=large,X=1.
$P(X = 0)_{small} = \frac{4}{8}$, $P(X = 1)_{large} = \frac{4}{8}$ We also denote fraction labelled T for Age=Child with y=1, and fraction labelled F for Age =Child with y=0 For Size,

$$P(y = 1|X = 1) = \frac{2}{4}, P(y = 1|X = 0) = \frac{3}{4}, P(y = 0|X = 1) = \frac{2}{4}, P(y = 0|X = 0) = \frac{1}{4}$$

$$MajorityError(y|X)_{size} = \frac{4}{8}min(\frac{3}{4}, \frac{1}{4}) + \frac{4}{8}min(\frac{2}{4}, \frac{2}{4})$$

$$MajorityError(y|X)_{size} = \frac{4}{8} * \frac{1}{4} + \frac{4}{8} * \frac{2}{4} = \frac{3}{8}$$

$$IG_{Size,Age=Child} = \frac{3}{8} - \frac{3}{8} = 0$$

We consider attribute Act next.
Lets denote Act = Stretch,X=0 and Act=Dip,X=1.
$P(X = 0)_{Stretch} = \frac{3}{8}$, $P(X = 1)_{Dip} = \frac{5}{8}$
For Act,

$$P(y = 1|X = 1) = \frac{3}{5}, P(y = 1|X = 0) = \frac{2}{3}, P(y = 0|X = 1) = \frac{2}{5}, P(y = 0|X = 0) = \frac{1}{3}$$

$$MajorityError(y|X)_{Act} = \frac{3}{8}min(\frac{2}{3}, \frac{1}{3}) + \frac{5}{8}min(\frac{3}{5}, \frac{2}{5})$$

$$MajorityError(y|X)_{Act} = \frac{3}{8}$$

$$IG_{Act,Age=Child} = \frac{3}{8} - \frac{3}{8} = 0$$

$$IG_{color,Age=Child} = 0$$

$$IG_{size,Age=Child} = 0$$

$$IG_{Act,Age=Child} = 0$$

Since, all have same IGs we pick Color to split the tree further on.

Color	Size	Act	Age	Inflated	Count
Blue	Small	Dip	Child	T	3
Blue	Large	Stretch	Child	F	1
Blue	Large	Dip	Child	F	1

Color	Size	Act	Age	Inflated	Count
Red	Small	Dip	Child	F	1
Red	Large	Stretch	Child	T	2

From the table we observe, that after splitting on Color =Blue, Size gives labels clearly and therefore we pick Size to split on.
On splitting on Size, for Small we get Inflated = T and Large gives Inflated = F. For Color = Red, we see that Size and Act both give labels clearly. We pick Size since it appears first in the table. Size = Small gives Inflated = F and Size =Large gives Inflated = T.
We finally get out full decision tree. It is described as below.

```
if age = adult:
  if size = small:
    if color = blue:
       if act = stretch:
        inflated = F
       if act = dip:
         inflated = T
    if color = red:
      inflated = F
  if size = large:
    if color = blue:
      inflated = F
    if color = red:
      inflated = T
if age = child:
  if color = blue:
    if size = small:
      inflated = T
    if size = large:
      inflated = F
  if color = red:
    if size = small:
      inflated = F
    if size = large:
      inflated = T
```

(c) No, ID3 does not give a globally optimal decision tree. ID3 follows a greedy heuristic approach, making a local greedy choice at each level. Thus, it generally gives local optima instead of global optima.It might also overfit the training data due to the presence of variance and noise and this might lead to the need for pruning.
Thus,ID3 produces small trees but not always the smallest possible trees due to its heuristic approach.

Q2. **Report:**Our dataset consists of names consisting of first,middle and last name alongwith a label(+/-).
Features: description of the features used to train the classifiers.

SGD with stumps: Stumps are Decision Trees with limited depth, generally of depth = 1. In this classifier, we don't use 390 simple features as described for other classifiers. Instead using stumps, we create a new feature set for the SGD classifier to be trained on.
We first get a list of 200 Decision tree stumps. A total of 200 features is obtained each of which is a prediction from a different decision tree stump obtained before. Thus our new feature matrix becomes n x 200.

SGD Classifier : We find features for every name, which represents one instance each. Every instance is represented by a list of 390 features (26 features for every character in the name, where name $\epsilon[first, middle, last]$) Thus across all instances say, of length n, we have a list of length n, with every item in the list, in this case every name, represented by a list of length 390. This gives us a feature matrix n x 390.
n depends on the length of the training file/folds. We cross validate using the 5 training folds as features and obtain the p_A and tr_A for all classifiers.

Decision Tree Classifier : We use the same feature matrix as the one mentioned above for SGD classifier to train a Decision Tree Classifier of full depth, depth = 4 and depth = 8.

Parameters: description of the parameters used for each classifier.
SGD with stumps:The decision tree stumps are trained using depth = 8 and criterion = 'entropy'. The SGD classifer is then trained on loss = 'log' using the feature set obtained using DT classifier.

SGD Classifier: The SGD classifier is trained using loss = 'log' on input data with learning rate = 'optimal'.Optional parameters alpha and tol are also used, to observe the relative performance.

Using pA as the performance criteria, we observe performance for different learning rates by varying alphas and error threshold by changing tol parameter. It is observed that by increasing alpha the pA increases. Therefore, as seen from the table below, we pick alpha=0.001 as it gives the best pA.

p_A	tr_A	alpha	tol
0.6571	0.8598	0.001	1e-3
0.6399	0.8510	0.0005	1e-3
0.6234	0.8385	0.0001	1e-3
0.62	0.8110	0.00001	1e-3
0.6514	0.8967	0.0001	1e-2
0.65	0.8857	0.0001	1e-3
0.645	0.8839	0.0001	1e-4
0.638	0.8981	0.0001	1e-5

Considering the above table, we also check pA for different error thresholds(tol). The default value is 1e-3.
It is observed that with decreasing tol(threshold) the performance decreases.

Therefore, for SGD Classifier we observe that a classifier with high value of alpha and high value of error threshold would give better performance as compared to those classifiers having low alpha and tol value.

Decision Tree Classifiers(depth = 4):The decision tree classifier is trained on criterion='entropy'. The depth is initialized to be 4. This restricts the tree depth to 4.

Decision Tree Classifiers(full depth): The decision tree classifier is trained on criterion='entropy'. The depth is initialized to be None. This keeps the tree depth unbounded.

Decision Tree Classifiers(depth = 8):The decision tree classifier is trained on criterion='entropy'. The depth is initialized to be 8. This restricts the tree depth to 8.

The ranking of all of the classifiers :

Algorithm	p_A	tr_A	Conf. Interval
SGD + Decision Stump Features	0.6428571421	0.9270714285	(0.711398045,0.5567305268)
Simple SGD	0.6268571428	0.8216414258	(0.730427126, 0.5143001678)
Decision Tree - Height 4	0.5971428571	0.6428571428	(0.6312081464, 0.5545061391)
Full Decision Tree	0.5942857142	1.0	(0.663611413, 0.5378171561)
Decision Tree - Height 8	0.5642857142	0.7425	(0.641902726, 0.4923829881)

From the table we observe the statistical significance for consecutive classifiers.
1)SGD-DT and Simple SGD: Since the p_A for SGD-DT(0.6428) lies in the confidence interval of Simple SGD. Thus, this pair is not statistical significant.
2)Simple SGD and DT(Height =4) : The p_A for Simple Sgd(0.6268) is contained within the confidence interval of DT(Height=4). Therefore, they are not statistically significant.
3)DT(Height=4) and DT(Full): The p_A for DT(Height=4) is contained in the confidence interval for Full DT, thus not statistically significant.
4)DT(Full) and DT(Height=8): The p_A for DT(Full) is contained within the confidence interval for DT(Height=8). Therefore, not statistically significant.

Thus,there are no consecutive pairs of classifiers with statistical significance.

Conclusion:
From the above evaluation, it can be seen that SGD with Decision Stump

Features gives the highest p_A, hence the best performance among the 5 classifiers discussed.The tr_A , training accuracy is also significantly higher than most classifiers, with the exception of Full Decision Tree.
Hence, SGD with stumps gives the best performance.

Comments: discussion of whether the results met expectations, tr_A versus p_A, and compare the results of the different algorithms.

Among the Decision Tree,SGD classifier and SGD-DT, I expected that SGD-DT would give a better performance since it uses a boosting technique by using a second model in this case SGD to improve the errors generated in the first model, in this case DT.Thus, it combines two classifiers with the expectation to get improved performance.
Decision trees are prone to overfitting.The tendency to overfit can be seen for the FullDT which has tr_A of 1.0. With increasing depth restriction the tr_A decreases as expected. The same however is not reflected for testing accuracy p_A among the trees. The Decision Tree(Height=4) performs the best with highest p_A among DTs followed by Full Decision Tree p_A being only slightly poorer than DT(4). DT(8) performs the poorest in terms of p_A.

Among SGD-DT and SGD, I expected SGD-DT to perform better and it does. The performances for default value of alpha = 0.0001 as shown in table ranks SGD-DT above SGD.However, by testing different values of alpha for SGD classifier,in this case the highest, alpha=0.001, the Simple SGD performs comparatively equally to SGD-DT.