1. Learning Decision Trees

  a. I wrote a python script to do both of these parts. The code is in script1.py. My ID3 tree is:

```
if Color = Purple:
    if Act = Stretch:
        if Age = Adult:
            class = T
        else:
            class = F
    else
        class = F
else:
    if Size = Small:
        class = T
    else:
        if Act = Stretch:
            if Age = Adult:
                class = T
            else:
                class = F
        else:
            class = F
```

  b. The code for this heuristic is in the same aforementioend script.py script. I actually ended up with the same tree as part [a.]. I designed it so that the MajorityError of an attribute is the minimum of the MajorityError of each class split. This resulted in the exact same tree as above. (See part [a.] if you want to see the tree.)

  c. Again, for both heuristics I got the same tree when training on the first 12 examples with maximum depth of 2. My trees for both were:

```
if Color = Purple:
    class = F //This was actually (if Act=Dip) => False, (else) => False,
             //when generated by my code, so I just simplified to this.
else:
    if Size = Small:
        class = T
```

```
    else:
        class = F
```

Testing on the remaining 4 instances, the two algorithms get 1/4, or 25% incorrect on the data. It makes sense that our accuracy on the training data went down by limiting the size of the tree. By doing this, our hypothesis becomes less expressive and begins to underfit the data. So when we test on the remaining fourth of our original dataset it seems only natural that it'd get at least 1 wrong.

2. Decision Trees as Features

   a. I chose not to change any features from the original features given. Also, for all of the problems below, I stuck with Weka to run the algorithms.

   b. My accuracies from the 5-fold cross-validation were : 340/400, 333/400, 339/400, and 328/400, giving an overall accuracy, $p_a$, of 83.34%. This yields a 99% confidence interval of [0.816,0.851].

   c. My accuracies from the 5-fold cross-validation were: 355/400, 351/400, 343/400, 354/400, and 360/400, with an average of $p_a = 88.15\%$. This gives a 99% confidence interval of [0.863,0.900].

   d. My accurances for the tree of maximum depth 4 from the 5-fold cross-validation were: 310/400, 310/400, 304/400, 316/400, and 311/400, with an average of $p_a = 77.55\%$. This gives a 99% confidence interval of [0.763,0.788].
   My accuracies for the tree of maximum depth 8 from the 5-fold cross-validation were: 333/400, 335/400, 324/400, 324/400, and 348/400, with an average of $p_a = 84.20\%$. This gives a 99% confidence interval of [0.801,0.883]

   e. My accuracies for the SGD trained on ID3 stumps from the 5-fold cross-validation were: 346/400, 346/400, 348/400, 340/400, 340/400, with an average of $p_a = 86.00\%$. This gives a 99% confidence interval of [0.849,0.871]

3. Evaluation Overall, I was pretty underwhelmed by my results. The relative frequency of negative examples in the sample data was about 71%, so even if I had some function, $f(x) = False$, it'd perform almost as well as my "trained" algorithms. However, to be fair, when I tested this "zero function" on the 5-fold cross-validation, it scored around 66% with a pretty tight confidence interval. So the trained algorithms *do* perform better (to a certain degree of confidence), it's just that they're not remarkably better.

The order of accuracies, from least $p_a$ to greatest is:

   1. ID3 with maximum depth of 4 : $p_a = 77.55\%$, CI : [0.763,0.788]

   2. Stochastic Gradient Descent : $p_a = 83.35\%$, CI : [0.816,0.851]

   3. ID3 with maximum depth of 8 : $p_a = 84.20\%$, CI : [0.801,0.883]

   4. SGD trained on ID3 features : $p_a = 86.00\%$, CI : [0.849,0.871]

5. ID3 with unbounded depth : $p_a = 88.15\%$, CI : [0.863,0.900]

However, SGD, and ID3 depth 8 have overlapping confidence intervals, as do ID3 depth 8/SGD trained on ID3 and SGD trained on ID3/ID3 with unbounded depth, so we can't *really* say that one is better than the other with confidence; it could've just been an "unlucky" sample from the population. However we can say with confidence that ID3 with unbounded depth is better than the base SGD, which is better than ID3 depth 4, which is better than $f(x) = False$.