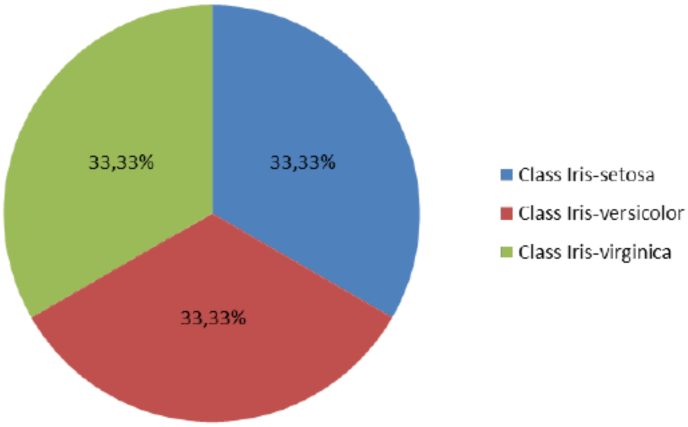
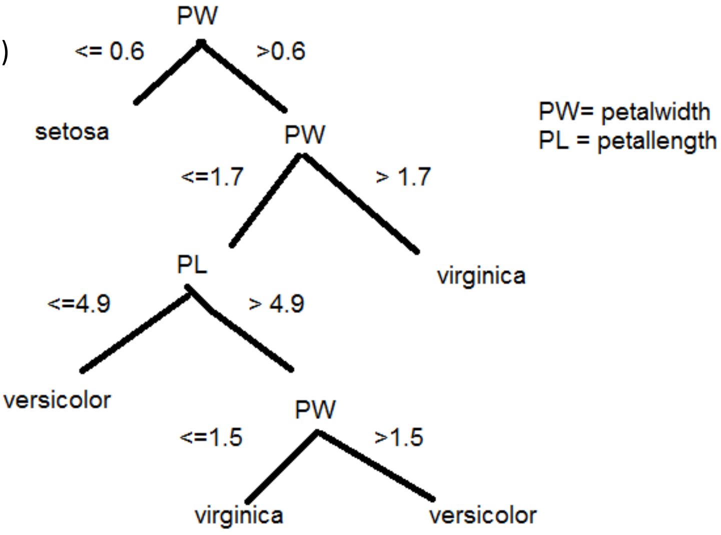
A little extra information about the Iris data set.

Classes: 3 (setosa, versicolor, virginica)
Attributes: 4
Instances: 150
Class balance:



Sample solution (98% accuracy on the entire data set)

```
petalwidth <= 0.6: Iris-setosa
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica
|   |   |   petalwidth > 1.5: Iris-versicolor
|   petalwidth > 1.7: Iris-virginica
```



(I know you aren’t evolving decision trees, this is just to give you an idea of a candidate solution)

From the solution above you will notice that a 98% accuracy can be obtained using only the features petal width and petal length. I used WEKA and some feature selection algorithm and these are the results I got:

```
Ranked attributes:
 1.418   petallength
 1.378   petalwidth
 0.698   sepallength
 0.376   sepalwidth
```

This means that petallength and petalwidth are better features than the two sepal ones, especially sepalwidth. This is just to give you a heads up that your GP trees could solve the problem using only two variables.

Some results from literature to give you an estimate on how your GP should perform:
These represent a range of results I found. They probably used 10-fold cross-validation too.

88.77 – 95.99[i]
93.30 – 95.30[ii]

[i] H. Liu, F. Hussain, C. L. Tan, and M. Dash, \Discretization: An enabling technique," Data mining and knowledge discovery, vol. 6, no. 4, pp. 393-423, 2002.

[ii] M. Hacibeyoglu, A. Arslan, and S. Kahramanli, \Improving classi_cation accuracy with discretization on datasets including continuous valued features," World Academy of Science, Engineering & Technology, 2011