1. **Use the following learning schemes to analyze the iris data (in iris.arff):**

|  |  |
| --- | --- |
| **ZeroR** | **- Weka.classifiers.ZeroR** |
| **OneR** | **- Weka.classifiers.OneR** |
| **Decision table** | **- Weka.classifiers.DecisionTable** |
| **C4.5** | **- Weka.classifiers.j48.J48** |

* **Do the decisions/model produced by the classifiers make sense to you? Why?**
* **How did each one of the methods perform? We will cover the evaluation techniques later in the class – for now you can choose common sense or one of the techniques that Weka presents with the model.**
* **Which method provided you with the most/least knowledge (incite into your data set/rules/patterns) and why?**

I will answer the three questions above at once as they are all inter-related.

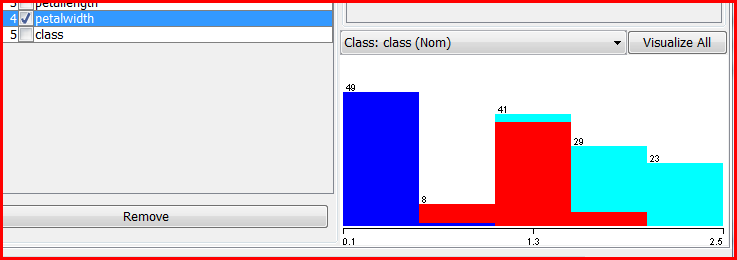
The Iris data in this file has 150 instances that evenly distribute into three groups (or types of iris): Iris-Setosa, Iris-Versicolor, and Iris-Virginica.

The decision/model produced by each of these four classifiers makes only sense and can only be fairly evaluated once we understand what each of them does, as well as their limitations and their purpose.

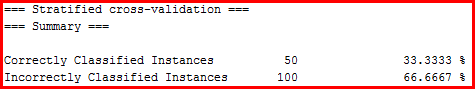
**ZeroR** only predicts the majority class correctly and is only useful for determining a baseline performance for other classification methods. That is why the results show in this case:

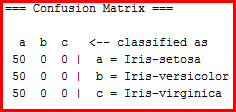
“*ZeroR predicts class value: Iris-setosa*”

This is why the Iris Setosa (blue area in chart below, based on Petal Width attribute) is the most ‘predictable’ class.

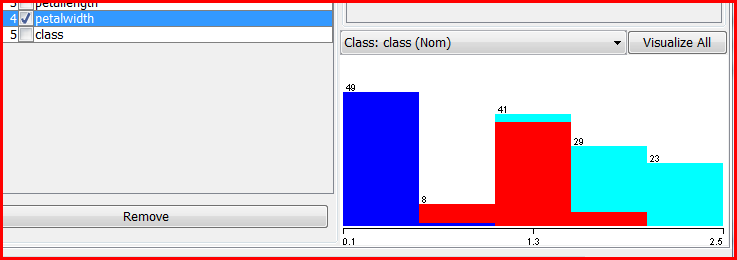


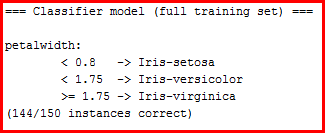
As ZeroR only predicts the majority class correctly, it can produce results that can be a bit confusing in a first reading, as shown, for instances, in the Classifier output -- where we can see that only the Iris Setosa instances get correctly classified and all other classifications fail (get classified as Iris Setosa too):





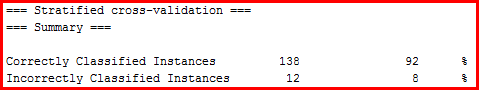
Again, this scheme is only valuable if it is considered as an exploratory tool and baseline performance for other classification methods.

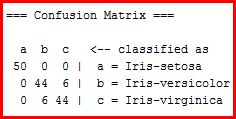
**OneR** is a simple but accurate classification algorithm that generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule". In the case of the Iris data set, as reflected in the first screen shot in my answers above , this single rule will be based on the Petal Width attribute:



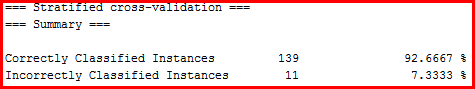
As explained by WEKA literature, OneR produces rules that, while are only slightly less accurate than state-of-the-art classification algorithms, are simple for humans to interpret. Our case study is a good proof of it, as we will see when we compare it to the next two schemes.

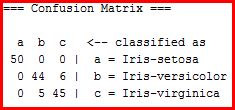
Below are two segments of the Classifier output so we can compare them with the results of the Decision Table and the C4.5. We can see how this simplistic rule, based on the Petal Width, is able to be a perfect classifier for the 50 instances of Iris-Setosa, while producing a big amount of correct classifications for the other two groups (Iris-Versicolor and Iris-Virginica).



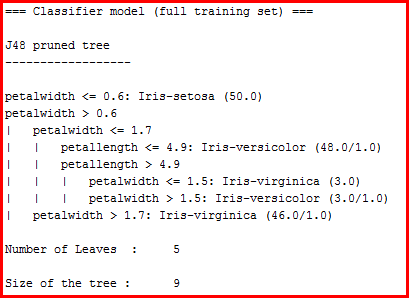


**Decision Table** is a classifier scheme for building and using a simple decision table majority classifier (in this case, Petal Width). As shown in screenshot below, compared to OneR, the Decision Table only adds one more correctly classified instance (one Iris-Virginica).

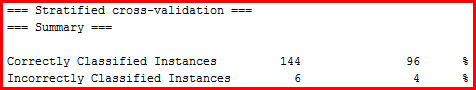




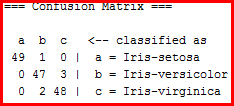
**C4.5** can be applied in WEKA as a J38 classifier, which allows us to generate a pruned or unpruned C4.5 decision tree. As we can see in the screen shot below (from the corresponding Classifier output), J48 comes out with a more complex, sophisticated scheme:



This scheme increases the overall prediction power:



Something interesting to note is how, spite of the increased prediction power of J48, it gets wrong one of the Iris-Setosas, which the other more simple schemes were able to get right. This could be the price to pay for the rule interaction or the pruning:



To finalize, I feel that I would rather use the benefits of all those schemes as a tool set: Although the C4.5 seems to be the most successful scheme and, if forced to make a choice of a single predictor tool among those 4 schemes for this exercise, I would pick this one, I would rather use them all. The ZeroR as an initial exploratory tool, OneR as a basic classification tool, both OneR and Decision Table to get supporting results to compare against the J48, and J48 as the most reliable classifier of the four in this case (although not by much, as mentioned above, from the Decision Table and the simple OneR).