**Group C Homework Group Credit Task**

**To be presented during the lecture on Tuesday, Sept. 19.**

Characterize and compare the following 3 histograms for the attribute sepal length that have been created for instances of Setosa, Virginica, and Versicolor Iris flowers. Also briefly assess the difficulties in separating the 3 types of Iris flowers using sepal length. You can assume in your analysis that the number of Setosas, Virginicas and Versicolors in the dataset that was analyzed is the same.



**Group D Homework Group Credit Task**

**To be presented during the lecture on Thursday, Sept. 21**

Compute the GINI-gain[[1]](#footnote-1) and information gain[[2]](#footnote-2) for the following decision tree split:

(10,5,5) (3,4,0)

(7,0,0)

 (0,1,5)

We assume we have a classification problem involving 3 classes C1, C2, C3. The above information can be interpreted as follows: Before the split there are 10 examples belonging to class C1, 5 examples of class C2, and 5 examples of C3 associated with the node. After a 3-way split 3 new nodes are introduced and 3 examples of class A and 4 examples of class B are associated with the first new node, 7 examples of class C1 are associated with the second new node, and 1 example of C2 and 5 examples of C3 are associated with the third new node.

1. (GINI before the split) minus (GINI after the split) [↑](#footnote-ref-1)
2. Entropy before the split minus entropy after the split. [↑](#footnote-ref-2)