Discussion of Learning Measures
# Introduction to Classification & Clustering Module

The "Introduction to Classification and Clustering" module has been used a number of times in classroom settings as part of an introduction to machine learning concepts. In each offering, students were presented with the "Start Up Survey" first to get a baseline starting point against which to measure learning. Then, module material was presented to students. Finally, a "Test Your Knowledge" survey quiz was given, providing a measure to compare student understanding before and after module presentation.

Classes in which measure was used:
- Spring 2015 - Machine Learning & Evolution class, 18 non-computer science majors
- Spring 2016 - Machine Learning & Evolution class, 23 non-computer science majors
- Fall 2016 - Machine Learning, 12 computer science majors, unfamiliar with ML concepts

Results

Start-up (pre) and Test Your Knowledge (post) surveys were given before and after presentation of the module material. Three nearly identical questions appear in both, and comparison measures significant learning. While it is expected that students knew little before presentation of the material, it is encouraging that when learning was measured at the start of the following class meeting students retained the understanding gained in the previous class meeting presentation, demonstrating a mastery of the concepts.
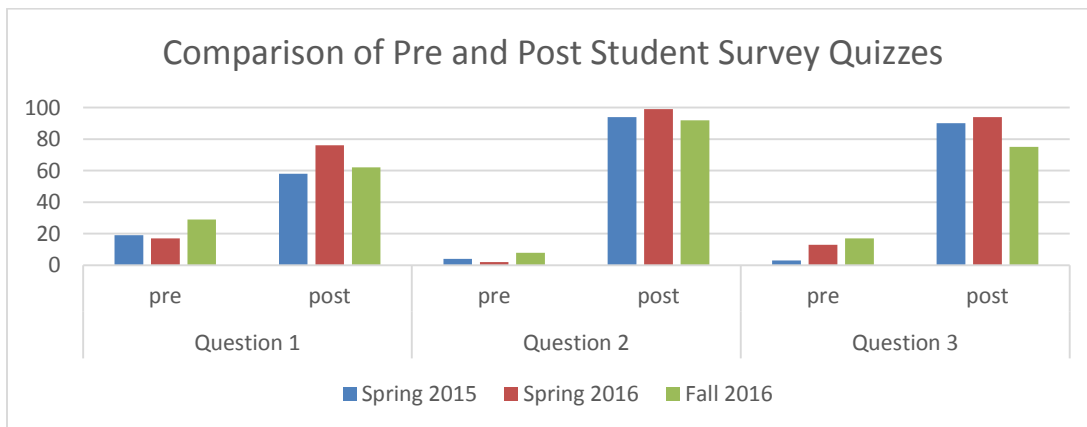
Specific observations where improvements can be made:

- In the Fall 2016 offering, on the Startup Survey some students described a specific pair of clusters rather than the process of clustering. On the Test Survey about a third of the students had improved to describe a process; the rest still described specific clusters.
- In the Fall 2016 offering, on question 2 of Test Your Knowledge, about half the students chose the GPS and paint matching foil. Neither of these requires learning; neither makes permanent changes to a computer's algorithm.

- In the Spring 2015 offering, students demonstrated slightly less increase in learning as measured by the questions than students in the Spring 2016 offering of the same class. This is likely due to slightly less time being available to cover the material due to a number of winter weather cancelations early in the Spring 2015 semester.

Comments

Timing

Generally, about one hour is spent on the module, including Startup Survey and Test Your Knowledge. This time could be significantly expanded or somewhat contracted depending on how much discussion there is about the activities and how much time is spent on "Make it Real."

Stand-up Activity

This has worked well for classes 12, 18 and 23 students and it typically only takes a few minutes for them to find a criterion that worked. Criterion have included gender, height, and even wearing a t-shirt with lettering (or not). Discussion always brings out other attributes that they had considered but rejected because the groups were not equal. The activity also works well to get the students involved and responsive, particularly if each student is required to say something about the experience. It is likely that classes larger than 30 students may not work well for the activity as described. Dividing larger classes up into smaller classes, space permitting, is suggested.

Hands-On Activity Comments

Classifying dogs and cats initially seems pretty trivial, and at first the only typical responses to "How did you approach it?" are along the lines of "They're dogs. Obviously." Some probing eventually does elicit specific features such as ear shape. The activity works best when it is emphasized that the images on the Group A and Group B pages are defining groups and that students should figure out how the images in the larger set matches those images.

In the clustering activities, an effective approach has been to ask each student to describe his or her clustering. This ensures that all students say something and seems to dramatically increase individual involvement. Performing the clustering activity into three groups typically elicits some very creative answers! The most recent winner for amusement value: "Whether I would like this one as a pet".

Make It Real Comments

In Fall 2016, discussing the emergency room example led into a discussion of some of the ethics of using a machine-learning-based system versus having a human make the decision. Points discussed included the error rate of each, the kind of errors humans and machine learning systems make, and the sensitivity of humans to edge cases that override the usual procedures.