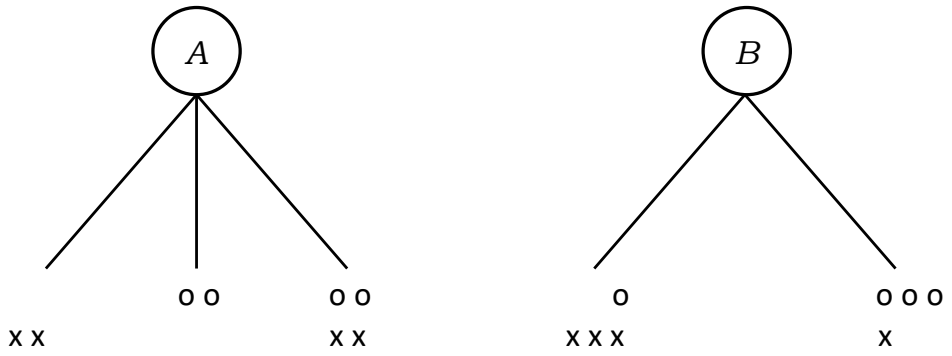
**Introduction to Machine Learning (COMP 170)**
**Assignment 01, Part 02 (40 points)**
***Due on Gradescope by 9:00 AM, Wednesday, 25 September 2019***

---

For this part of the assignment, you will answer a few technical and conceptual questions about the decision tree algorithm, and the two heuristics used in your implementation of it. Your answers should be typed up and submitted in PDF form to the Gradescope site for the course. See the end of this document for format requirements.

1. (*20 pts.*) In the following diagram, we see a small data-set consisting of two output classes (marked by $\times$ and $\circ$, respectively), along with the results of splitting the data according to the two properties $A$ (with possible values $a_1$, $a_2$, $a_3$) and $B$ (with possible values $b_1$, $b_2$):




   (a) (*15*) For each of the two heuristics employed in the decision tree algorithm (counting and information-based), work out which of the features, $A$ or $B$, would be chosen by that heuristic. In each case, show all work, including the calculations used to make the determination. You can assume that in the case of ties, each heuristic is used with a random tie-breaking procedure.

   (b) (*5*) What conclusions can you draw from these results? That is, what does this mean about the heuristics, and about the tree-building algorithm itself? (Your answer should consist of a few sentences, minimum, of reflection and analysis.)

2. (*10 pts.*) In the sample performance shown for the decision-tree algorithm (slides of 11 September, slide 28), we can see that the classification accuracy of the algorithm is not a monotonic function of training set size—that is, sometimes the performance goes down, even as more training data is added. You may or may not have seen similar performance in runs of your own program on the mushroom data-set.

   Why does this happen? That is, what could cause the algorithm to actually do worse when supplied with more information. Your answer should be as clear as possible, explaining what features of the algorithm, and what scenarios, exactly, can cause this to occur (for any given heuristic). *If you consult other sources in your thinking about this question, remember to cite those sources in your response.*

3. (*10 pts.*) Given the behavior noted in the previous question, we can say that the decision-tree algorithm is *sensitive* to the details of its training set. As a result, many modern applications of decision trees use variations on the base algorithm, often generating multiple distinct trees. Explain at least one of these approaches, and how it works exactly. Be sure to highlight ways in which it is meant to improve over the base algorithm. *If you consult other sources in your thinking about this question, remember to cite those sources in your response.*

---

**Format requirements**: work for COMP 135 should correspond to the following guidelines:

- Work must be in type-written format, with any diagrams rendered using software to produce professional-looking results. No hand-written or hand-drawn work will be graded.
- Work must be submitted in PDF format to Gradescope. When submitting, the system will ask you to indicate the page on which each answer appears; make sure that you do so correctly.

You can find links to information about using LaTeX to produce type-written mathematical work* on the class website:

http://www.cs.tufts.edu/comp/135/resources/

*LaTeX was used to produce this document.