Due Date: Monday, December 4, 2017 at 11:59pm.

Submission Instructions: All assignments are to be submitted through LEARN, in the Dropbox labelled *Assignment 5 Submissions* in the Assignment 5 folder. Late assignments will be accepted up until December 6 at 11:59pm. Please read the course policy on assignments submitted after the official due date. *No assignments will be accepted, for any reason, after 11:59pm on December 6.*

Lead TA: Mike Schaekermann (mschaeke@uwaterloo.ca). Office hours are Fridays from 11:00-12:00 in DC2306C (AI Lab).

Announcements: The following exercises are to be done individually.

isn't it like A* or somehow like this?

Question 1. Decision Tree for Equine Colic Diagnosis [40 points]

Equine colic is one of the leading causes of death in adult horses. However, if diagnosed early enough, it is often surgically curable. Using the language of your choice, write a decision tree algorithm that will learn to diagnose whether a patient is healthy or has colic. Use the **horseTrain** file, (found in the Assignment 5 folder) to train the decision tree. Each training instance has 16 numeric attributes (features) and a classification, all separated by commas. The attributes correspond to the following measurements made from each patient at admission to the clinic.

1. K
2. Na
3. CL
4. HCO_3
5. Endotoxin
6. Aniongap
7. PLA2
8. SDH
9. GLDH
10. TPP
11. Breath rate
12. PCV

13. Pulse rate
14. Fibrinogen
15. Dimer
16. FibPerDim

In the decision tree, use only binary tests, *i.e.* each node should test whether a particular attribute has a value greater or smaller than a threshold. In deciding which attribute to test at any point, use the information gain metric. Set the node test threshold for each potential attribute using this same metric *i.e.* at each point, see all the values that exist for a particular attribute in the remaining instances, order those values, and try threshold values that are (half way) between successive attribute values. Use the threshold value that gives the highest information gain. Allow the same attribute to be tested again later in the tree (with a different threshold). This means that along a path from the root to a leaf, the same attribute might be tested multiple times. After learning the decision tree, use the **horseTest** file to test the generalization accuracy of the tree.

Submit the following

1. (5 points) Your code.
2. (20 points) Picture of the decision tree. Hand drawn and then scanned is fine.
3. (5 points) Answer the question "How many of the training instances does the tree classify correctly?"
4. (5 points) Answer to the question "How many of the test instances does the tree classify correctly?"
5. (5 points) Description of how you used the information metric.

Format of data files

We are providing you with data from an actual vet clinic. Each line in a file is one instance. The first 16 numbers are the values for the 16 attributes listed above. The last entry on a line is whether the horse was healthy or not. You are allowed to reformat the data files in what ever way you want so as to make reading them easier for your program.

Question 2 [25 points]

In this question you will read a research paper and answer questions related to it. The goal of this assignment is to deepen your understanding of recent developments in artificial intelligence and to give you a chance to practice the skills required when reading advanced technical material.

1. Read the instructions in the file "How to Read a Research Paper" which is located in the Assignment 5 directory.

2. Chose **ONE** of the following papers and read it carefully

   - Fei Fang, Thanh H. Nguyen, Rob Pickles, Wai Y. Lam, Gopalasamy R. Clements, Bo An, Amandeep Singh, Milind Tambe, Andrew Lemieux, Deploying PAWS: Field optimization of the protection assistant for wildlife security, *IAAI-16: The Twenty-Eighth Annual Conference on Innovative Applications of Artificial Intelligence*, 2016.
   - Percy Liang, Learning executable semantic parsers for natural language understanding, *Communications of the ACM*, 59 (9), pp. 68–76, 2016.
   - David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis, Mastering the game of Go without human knowledge, *Nature*, pp. 354-359, 2017.

   These papers are found in the Assignment 5 directory.

3. For the paper you have chosen, answer the following questions:

   (a) What are the motivations for this work? (5 pts)

   (b) What is the proposed solution? (5 pts)

   (c) What is the evaluation of the proposed solution? (5 pts)

   (d) What are the contributions? (5 pts)

   (e) What are future directions for this research? (5 pts)

   For each question, you are expected to write at least one paragraph.