COMP 7745/8745: Machine Learning

INSTRUCTOR: Deepak Venugopal

SPRING 2019: HOMEWORK 1

DUE DATES: FEBRUARY 19, 2019 IN CLASS/ DROPBOX IN ELEARN

1. Can the following functions be represented using decision trees? If your answer is yes, draw the corresponding tree, if your answer is no, briefly state why. (12 points)

   - $A \wedge \neg B$
   - A XOR B
   - $A \vee (B \wedge C)$

2. Can the following functions be represented using perceptrons. If your answer is yes, compute the weight vector for the perceptron such that it can classify all instances of the functions correctly. Give a one line justification if your answer is No. (12 points)

   - $A \vee B$
   - $\neg A \vee B$
   - A XOR B

3. Suppose we have a classifier that classifies if an image contains a Human face or not. Suppose we have 100 images, 50 of which contain human faces. If our classifier accurately classifies that 30 images contains human faces, but at the same time wrongly classifies that 30 images contains human faces. What is the precision and recall of the classifier. If we change the classifier to detect more images as those that contain human faces, which is more likely to reduce, precision or recall? (10 points)

4. For each of the following, answer briefly (no formal math needed explain intuitively). (30 points)

   (a) Given a training dataset with N features, is the number of nodes in any decision tree learned from this dataset guaranteed to be lesser than or equal to N? Why or why not? Briefly explain. (10 points)

   (b) Suppose we have a linearly separable dataset, and we divide the data into training and validation sets. Will a perceptron learned on the training dataset (assuming gradient decent works perfectly well) be guaranteed to have i) 0 error on the training dataset ii) 0 error on the validation dataset. Briefly explain. (10 points)

   (c) Suppose your boss asks you design a ML algorithm for real-time prediction. Specifically, the requirement is that the ML algorithm needs to preform predictions very quickly. Can decision trees be used for such an application? Briefly explain your reasoning. (5 points)

(d) Given a dataset where the dataset is not linearly-separable, and each of the features have continuous values, which of the following algorithms is more ideally suited a) perceptron b) decision-trees c) neural-networks. Why? (5 points)

5. Run the ID3 algorithm (manually) for the following dataset to classify whether students like a restaurant or not. (10 points)

Price	Fast	Distance	Like
Low	No	Near	Yes
Low	Yes	Far	Yes
High	No	Near	No
High	Yes	Far	No

6. A 2-layered neural network where each perceptron is a linear unit (no thresholding) has the same expresiveness as a 2-layered neural network where each perceptron is a sigmoid unit. Briefly explain whether the statement is true or false. (Slightly more challenging question. Hint: think of a simple example NN with only linear units and check what happens) (12 points)

7. Experiments (14 points) You will predict survivors of the Titanic disaster (titanic.csv) using the J48 algorithm. The dataset is included as titanic-1.csv. The features are the $passenger-ticket-class$, sex, age, $number-of-siblings-in-ship$, $number-of-parents-in-ship$, $embarking-port$ and the label indicates whether they survived or not. Draw a graph that shows avg. precision vs avg. recall (for 5-fold cross validation), for varying values of the pruning confidence parameter (e.g. 0.1,0.25,0.5 and 1). Did pruning greatly affect performance?

You will predict the wine quality using the wines dataset using neural networks (implemented as multilayer perceptrons in Weka under functions). For Neural networks, vary the learning rate (0.01,0.1,0.2) and the number of hidden layers (1,2). Create a table with these results. Which combination of learning rates and hidden layers has the best performance in terms of average F-1 score? Modify the validationsetsize parameter and check if this improves/degrades your results.