**November 5, 2018 Review for COSC 4355 Midtem2 Exam**

**~~1) Hierarchical Clustering~~**

~~Hierarchical Clustering algorithm creates dendrograms; what is a dendogram? How are the clustering results K-means creates differerent from those of hierarchichal clustering algorithms?~~

**~~2) DBSCAN~~**

**~~a)~~** ~~Assume I run DBSCAN with MinPoints=6 and epsilon=0.1 for a dataset and I obtain 4 clusters and 5% of the objects in the dataset are classified as outliers. Now I run DBSCAN with MinPoints=5 and epsilon=0.1. How do expect the clustering results to change?~~

**~~b)~~** ~~Assume we have two core points o and v that are within each other’s radius—will o and v belong to the same cluster? Now assume, that we have a border point b within a radius of a core point u—will b and u always belong to the same cluster? Give reasons for you answer! [3]~~

**3) Classification**

1. Compute the GINI-gain[[1]](#footnote-1) for the following decision tree split (just giving the formula is fine!)[3]:

(12,4,6) (3,3,0)

(9,1,0)

 (0,0,6)

1. Assume there are 5 classes; Compute the entropy of the following class distribution: (1/2,1/4.1/8,1/8,0), giving the exact number not only the formula! [2]

c) What is overfitting? What are the characteristics of overfitting? What can be done in the context of decision trees to battle overfitting?

**d)**The following dataset is given (depicted below) with A being a continuous attribute and GINI is used as the evaluation function. What root test would be generated by the decision tree induction algorithm? What is the gain (equation 4.6 page 160 textbook) of the root test you chose? Please justify your answer![6]

Root test: A >=

|  |  |
| --- | --- |
| A | Class |
| 0.22 | 0 |
| 0.22 | 0 |
| 0.31 | 0 |
| 0.33 | 1 |
| 0.33 | 1 |
| 0.41 | 0 |
| 0.41 | 1 |

**Possible slits**

**A≤0.22: (0,2); (3,2)**

**A≤0.31: (0,3); (3,1)**

**A≤0.33: (2,3); (1,1)**

as A≤0.31has a purity of 100%/75% which is much higher than the purity of the other splits, this split will be selected.

e)Most decision tree tools use gain ratio and not GINI or information gain in their decision tree induction algorithm. Why? [3]

74. **Computing Entropy using R [11]**

Write a function *H* in R[[2]](#footnote-2), whose input is a vector of class proportions of arbitrary length[[3]](#footnote-3) called v (v contains O and positive numbers whose sum is exactly one) and returns the entropy of for v; e.g.

v<-c(0.5, 0.25, 0.25, 0)

H(v)

*would return:* 0.5\*log2(2) + 2\*1/4\*log2(4) + 0=1.5

Remark: Values of 0 in the input vector do not make any contributions to the overall entropy—their contribution is 0; therefore, make sure when you write the code of the H function that you do not compute 0\*log2(0) as this will return NA[[4]](#footnote-4).

**5) SVMs [9]**

a) What are the characteristics of hyperplanes that support vector machines learn from a training set? [3]

b) The soft margin support vector machine solves the following optimization problem:



What does the second term minimize? Depict all non-zero ξi in the figure below! What is the advantage of the soft margin approach over the linear SVM approach? [5]

All other points

have **ξi** values

of 0!

width

width

c) Referring to the figure above, explain how examples are classified by SVMs! What is the relationship between ξi and example i being classified correctly? [4]

d) Assume you use an ensemble approach. What properties should base classifiers have to obtain a highly accurate ensemble classifier? [3]

6) Neural Networks

How are activation functions used in neural network computations? What is neural network learning all about? Give a brief sketch how multi-layer neural networks learn models.

7) KNN

a) k-NN employ a lazy approach to learning models from training examples. What does this mean? What disadvantages you see with k-NN’s lazy learning approach? [3+1 extra point}

b) What can be said about the number and shape of decision boundaries of a k-NN classifier? [3]

The shape consist of lines formed of edges of convex polygons of the Vonoroi[[5]](#footnote-5).

c) How does a 3-nearest neighbor classify an example?

1. (GINI before the split) minus (GINI after the split) [↑](#footnote-ref-1)
2. You will need to write your own code; calling a function in an R-package which computes entropy will not get much credit! [↑](#footnote-ref-2)
3. You can use the length function to determine how many numbers v contains. [↑](#footnote-ref-3)
4. Moreover, in R, log (8,2) computes log2(8). [↑](#footnote-ref-4)
5. Its number is only bounded by the number of examples in the training set. [↑](#footnote-ref-5)