**November 6, 2019 Review for COSC 4355 Midtem2 Exam**

**1) Classification**

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]

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

**c)** 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.

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

**2) 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]

3) Neural Networks

a) 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.

b) What is the purpose of the backpropagation algorithms

4) 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[[1]](#footnote-1) tessalation.

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

5) Similarity Assessment [9]

Design a distance function to assess the similarity of electricity company customers; each customer is characterized by the following attributes:

1. Ssn
2. Oph (“*on-time payment history*”) which is ordinal attribute with values ‘excellent’, ‘good’, ‘medium’, and ‘poor’.
3. Power-used (which is a real number with mean 2000, standard deviation is 1000, its maximum is 10000 and minimum 100)
4. Country\_of\_Citizenship is a nominal attribute

Assume that the attributes Oph and Power-used are of major importance and the attribute Country\_of\_Citizenship is of a minor importance when assessing the similarity between customers. Using your distance function compute the distance between the following 2 customers: c1=(111111111, ‘excellent’, 2000, ‘Peru’) and c2=(222222222, ‘good’,2100, ‘France’)!

We convert the Oph rating values ‘excellent’, ‘good, ‘medium’, and ‘poor’ to 3:0 and then divide by the numbers by the number of values minus 1; that is by 3 in this case**;** finally we compute the distancesby applying the L-1 norm to the mapped values of the attribute Oph.

Normalize Power-used using Z-score and find distance by using L-1 norm

dCountry\_of\_Citizenship(a,b):= if a=b then 0 else 1

Assign weights 0.4 to Oph, 0.4 to Power-used and 0.2 to Country\_of\_Citizenship

Now:

**d(u,v) = 0.4\*|(u.PowerUsed – v.PowerUsed)/1000| + 0.4\*|φ(u.Oph) – φ(v.Oph)|/3 + 0.2\*dCountry\_of\_Citizenship(u.Country\_of\_Citizenship, v.Country\_of\_Citizenship)**

2 customers:

c1=(111111111, ‘excellent’, 2000, ‘Peru’) and c2=(222222222, ‘good’,2100, ‘France’)

= 0.4|(3 - 2 )/3| + 0.4\*|(2000 - 2100)/1000| + 0.2\*1= 0.133 + 0.04 + 0.2 = 0.373

If distance functions do not make much sense give 2 points or less

Distance functions are not defined properly [-5]

One error [-3]; two errors [-6]; more than 2 errors at most 1 point!

Only function is mentioned and no explanation: -4

Not calculated the distance between 2 customers: -2

6) What is the main challenge of ensemble learning?

To find a set of classifiers that make different kind of errors

7) Assume we have a datasets with classes C1, C2, and C3 and symbolic Attributes A1, A2, and A3? How does a naïve Bayes classifier determine the class label for an example with attribute values X={A1=’blue’,A2=’long’,A3=’serious’)?

1. Its number is only bounded by the number of examples in the training set. [↑](#footnote-ref-1)