CS585 Final exam

2016-12-08

Duration: 1 hour

Last Name: _____

First Name: _____

Student ID: _____

Email: _____

Question	Your score	Max score
1		2
2		2
3		2
4		6
5		4
6		4
7		3
8		3
9		2
10		2
Bonus		2
Total		32

Note - the exam is closed book/notes/web/neighbors(!), and ‘open mind’ :)

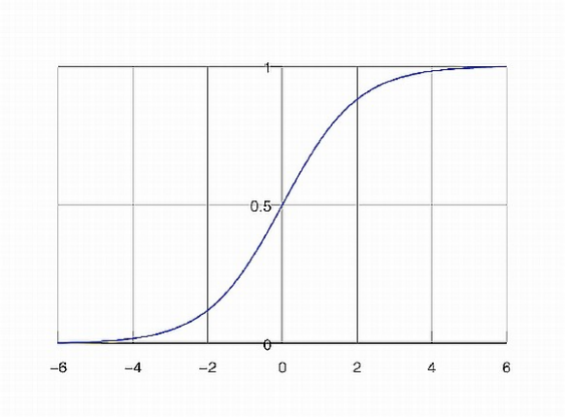
Good luck!
- Saty

Q1 (2 points). What role do minimum bounding rectangles (MBRs) play, in spatial query processing (how are they used/helpful)?

Q2 (2 points). 'Ensemble methods' are often used in machine learning – what is the single biggest benefit of using this technique?

Q3 (2 points). In WEKA, a native (custom) file format is used to read a table. Name the format, and provide a very small example.

Q4. (3+3=6 points). The 'sigmoid' function/curve shown below, is useful in at least two techniques of Machine Learning. What are the two techniques, and briefly, how is the curve used in each?



Q5 (2+2 = 4 points). Fraudulent credit card purchase detection relies on using a binary (yes/no) classifier to analyze card transactions. Name two algorithms that could be used for this purpose (we covered four), explain very briefly how each works.

Q6 (1*4=4 points). A very straightforward question – name (the) 4 types of NoSQL DBs, and provide an example (an open source or commercial implementation) of each.

Q7 (1+2 = 3 points). The MapReduce algorithm has a step between Map and Reduce - what is the step?

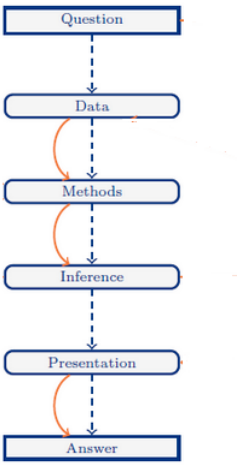
Briefly explain, with a diagram, how MapReduce works.

Q8 (2+1 = 3 points).
a. What are two capabilities that MapReduce v2 (YARN) provides, that v1 does not?

b. There is an architecture that can serve an alternative in some cases to MapReduce, for Big Data processing – what is it? Just naming it is adequate.

Q9 (1+1 = 2 points).

a. What does the following diagram summarize?



b. Name 4 items you would list under 'Methods' shown above.

Q10 (1+1 = 2 points). Look at the code below.
a. In what language is the code written?

b. What does it do (please be specific)? Examining the code and reading the comments should easily lead you to the answer :)

```
install.packages('neuralnet')
library("neuralnet")
#Generate 50 random numbers uniformly distributed between 0 and 100
#And store them as a dataframe
traininginput <-  as.data.frame(runif(50, min=0, max=100))
trainingoutput <- sqrt(traininginput)
#Column bind the data into one variable
trainingdata <- cbind(traininginput,trainingoutput)
colnames(trainingdata) <- c("Input","Output")
#Train the neural network
#Going to have 10 hidden layers
#Threshold is a numeric value specifying the threshold for the partial
#derivatives of the error function as stopping criteria.
net.sqrt <- neuralnet(Output~Input,trainingdata, hidden=10, threshold=0.01)
print(net.sqrt)
#Plot the neural network
plot(net.sqrt)
#Test the neural network on some training data
testdata <- as.data.frame((1:10)^2) #Generate some squared numbers
net.results <- compute(net.sqrt, testdata) #Run them through the neural network
#Lets see the results
print(net.results$net.result)
#Lets display a better version of the results
cleanoutput <- cbind(testdata,sqrt(testdata),as.data.frame(net.results$net.result))
colnames(cleanoutput) <- c("Input","Expected Output","Neural Net Output")
print(cleanoutput)
```

Bonus (2 points). Consider the four cards shown below - each has a letter on one side, and a number on the reverse. Now consider this statement: "Every card with a vowel on one side has an even number on the other side." How many cards minimum would you need to turn over, to find out if the above statement is true or false? You need to name which card[s] you will flip, and why. Note that the answer is one of 1,2,3 or 4 :)