Brief Homework 3
Computer Science
Fall 2016
B565

Professor Dalkilic

September 28, 2016

Directions

Please follow the syllabus guidelines in turning in your homework. I will provide the LaTeX of this document too. You may use it or create one of your own. This homework should be started quickly. Sometimes there are natural questions arising from code. Within a week, AIs can contact students to examine code; students must meet within three days. The session will last no longer than 5 minutes. If the code does not work, the grade for the program may be reduced. Lastly, source code cannot be modified post due date.

Questions

1. In `R` load the iris data. Issue the following commands at the prompt:

```
> data(iris)
> pairs(iris[1:4],main="Iris Data", pch=19, col=as.numeric(iris$Species)+1)
> iris.pca <- prcomp(log(iris[,1:4]), scale=TRUE, center=TRUE)
> summary(iris.pca)
> iris.pca
> screeplot(iris.pca, type="lines", col="3")
>> iris.pca$sdev^2
```

   (a) What species is red, blue, and green?

   (b) What are the principle components?

   (c) There are three methods to pick the set of principle components: (1) In the plot where the curve *bends*; (2) Add the percentage variance until total 75% is reached (70-90%) (3) Use the components whose variance is at least one. Show the components selected in the iris data if each of these is used.

   (d) What are the first two unit eigenvectors? What does *unit* mean?

   (e) How does the PCA differ when no log transform is done of the data?

2. Assume $X = \{(a,2),(b,2),(c,16)\}, Y = \{(a,4),(b,1),(c,15)\}, Z = \{(a,10),(b,4),(c,4)\}$ are multisets. Calculate the entropy on each. Does is make sense to compare entropy between X, Y, X, Z, and Y, Z? Assume you have another multiset W and its entropy is 0. How many elements does it have? Assume its entropy is the same as X. How many elements does it have? What's the primary difference between multisets and probability distributions?

3. Consider the following data:

$$\Delta \quad = \quad \{((a,a,a),5),((b,b,a),10),((a,a,b),5),((b,b,a),10)\} \tag{1}$$

   Assume we name the attributes A_1, A_2, A_3 and label L. Create a relation instance $\mathbf{r}$ that describes Δ. Show explicitly which attribute is best to split in using Information Gain. Split on that attribute, then show which attributes are best again. Write the tree as a rule set. What is the label of (a,b,a)?

4. ID3/C4.5 is a greedy algorithm. What is the implication for the kind of trees ID3 produces?

5. (True or False) Error in the training set is usually more or less than the true error of the real function?

6. Assume you have the dissimilarity matrix Q of data A,B,C,D.

	A	B	C	D
A	0			
B	1	0		
C	4	8	0	
D	5	2	2	0

   Using this ultrametric distance:

$$Q(x,(yz))) \quad = \quad \min\{Q(x,y), Q(x,z)\}$$

   Show a tree that results when using hierarchical alggomerative clustering. Show how Q changes at each iteration.

7. (True or False) For clustering that has less than 1000 data points, enumeration of all possible clusters is feasible.

8. (True or False) The run-time of agglomerative clustering is $O(n^2)$.

9. Assume a data set Δ for classification has two attributes A_1, A_2 that are binary, $dom(A_1) = dom(A_2) = \{0,1\}$ and label $L = \{0,1\}$ with a binary outcome. The information gain for *both* A_1, A_2 are identical and $\mathcal{H}(\Delta_{[L]}) = k$ for $k > 0$. Fill-in the entries below that satisfy these constraints.

A_1	A_2	L	Count
$\vdots$	$\vdots$	$\vdots$	$\vdots$

10. Read Quinlan's, "Induction of Decision Trees," Machine Learning 1:81-106 (1986), Kluwer Academic Publishers. In no more than four paragraphs, discuss what he writes about Noise and Unknown Attribute Values for ID3. How will this impact your use of ID3/C4.5?

11. Let $\Omega = \mathcal{Z}$. We define a random variable X as:

$$p_X(x) \quad = \quad \begin{cases} 1/9, & -4 \leq x \leq 4 \\ 0, & o.w. \end{cases} \tag{2}$$

    (a) Calculate E[X]

    (b) Let $Y = |X|$. Calculate E[Y]

    (c) Calculate Var[X]

12. Let Ω be the throw of a fair die twice. Let A be the event a 3 on each individual throw was observed or the sum is at least 5. Let B be the event that the difference between the two throws is exactly one. Are A, B independent? What is the probability of B given A? What is the probability of A given B?

13. Consider the a random variable X with probability distribution:

$$f_X(x) \quad = \quad \begin{cases} \frac{1}{2}x^{-\frac{1}{2}}, & 0 < x \leq 1 \\ 0, & o.w. \end{cases} \tag{3}$$

    Show this is a probability distribution even though, as x approaches zero, the values become very large. Find the expectation of X.

What to Turn-in

The *pdf of the written answers to this document.