

Homework 5

CS540 section 2

due 8th December 2005

Problem 1. Perceptrons (9 points)

Using as few perceptrons as possible, create a neural network that captures each of the following propositional logic sentences. Use a simple step-function at each perceptron. Label weights and the activation threshold on each perceptron in the network. To keep things simple, use only integer values for weights.

(a) $A \mid B$ This is known as the sheffer stroke. Its truth table is the following:

A	B	$A \mid B$
1	1	0
1	0	0
0	1	0
0	0	1

(b) $A \Rightarrow B$

(c) $A \Leftrightarrow B$

Problem 2. Neural Networks (13 points)

Consider a learning task that has three input features (A, B, and C) and one output (D). You decide to learn this task using neural networks. So you create a topology containing *two* ($H1$ and $H2$) hidden units, whose activation function is the sigmoid (see Figure 20.16b). You fully connect each input unit to every hidden unit and you connect each hidden unit to the output unit.

(a) Draw this neural network and initialize all the weights into and out of $H1$ to 0.3 and all of the weights into and out of $H2$ to -0.1. Initialize the "bias" (threshold) of each hidden and output unit to 0.2. (In practice, these weights would be set to small, randomly chosen, numbers, but for grading simplicity we're specifying the initial values).

(b) Consider the following training example and show how the back-propagation algorithm (Figure 20.25) would change your network's weights and biases. Use a learning rate of 0.3.

$$A = 0.3, \quad B = 0.8, \quad C = 0.1, \quad D = 1$$

Problem 3. Probability (8 points)

The cat is out of the bag. Here is how I plan on deciding who, and how many will be passing my class:

	P		$\sim P$	
	M	$\sim M$	M	$\sim M$
C	0.75	0.02	0.05	0.015
$\sim C$	0.08	0.005	0.07	0.01

Let P mean “pass the course”, C mean “CS major”, and M mean “passed midterm”.

(a) What is $P(P|C)$? Show your work.

(b) What is $P(P|C,M)$? Show your work.

(c) What is $P(P)$? Show your work.

(d) What is $P(C,M)$? Show your work.

(e) Are CS and Midterm independent? Show your work.

(f) Are CS and Midterm conditionally independent given Pass? Show your work.

* -- Disclaimer: no students were harmed in making this question and any resemblance to life in actuality is purely coincidental.

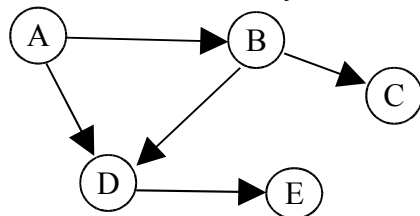
Problem 4. Bayesian Networks (18 points)

Using the Bayesian Network below to answer the following questions.

(a) $P(E|B,C)$ using inference by enumeration. Show your work.

(b) $P(D|E,A)$ using variable elimination. Show your work.

(c) For part (b) what additional node, if added to the evidence variables, would simplify the computation the most? Why?



The CPT's are as follows:

$$P(A)=0.8$$

$$P(B|A)=0.2$$

$$P(B|\sim A)=0.9$$

$$P(C|B)=0.95$$

$$P(C|\sim B)=0.50$$

$$P(D|A,B)=0.85$$

$$P(D|A,\sim B)=0.25$$

$$P(D|\sim A,B)=0.15$$

$$P(D|\sim A,\sim B)=0.05$$

$$P(E|D)=0.60$$

$$P(E|\sim D)=0.50$$

Problem 5. Clustering (12 points)

K-means clustering tries to minimize the distortion $\sum (x_i - c_i)^2$ where c_i is the centroid (mean) of the cluster that x_i is in.

(a) [6] Given the one-dimensional dataset $\{1, 2, 3, 5, 7\}$, and $K=2$ clusters whose initial centers are $c_1=0, c_2=9$, run K-means clustering by hand. Show

- i) the final cluster centers
- ii) the points in the two clusters respectively
- iii) the distortion

(b) [4] In general (not limited to the dataset in a) if we move a point x from cluster 1 to cluster 2, how would the center c_1, c_2 of the clusters change? Express the new centers c_1^{new}, c_2^{new} using only the old centers c_1, c_2 , the number of points n_1, n_2 in the old clusters, and x .

(c) [2] Let's go back to the dataset in a), if we move the point '3' to the other cluster, what is the new distortion?

Problem 6 (Programming-use Portion) Weka (20 points)

the purpose of this problem is to help you gain familiarity with one tool that is publicly available for performing many classification tasks. The program is called *Weka* and can be found at the website <http://www.cs.waikato.ac.nz/ml/weka/>. Documentation on how to use weka can be found on that site, but hopefully, with the in-class demo, the interface is clear enough for you to perform the tasks for this problem. In addition to the website, I have made the weka-3-4-6.zip file download-able from the course homepage. Download and unzip the file. Change to the weak-3-4-6 directory and type “java -jar weak.jar” to start up the program. Everything you will be doing is done inside the explorer.

UCI machine learning repository (<http://www.ics.uci.edu/~mllearn/MLRepository.html>) has a dataset for pen-based recognition of handwritten digits. The dataset was divided into a training and testing set as follows: “We create a digit database by collecting 250 samples from 44 writers. The samples written by 30 writers are used for training, cross-validation and writer dependent testing, and the digits written by the other 14 are used for writer independent testing.” I have provided a modified version of the dataset split into training and testing sets on the course homepage. Download the training and testing files and place them inside the data directory under weak-3-4-6.

After you start up weak and open the explorer, go to the pre-process tab and open the training file. Then switch to the classifier tab to perform tasks (a) and (b) below.

Tasks:

- (a) Using the training set construct a naïve bayes classifier and classify the examples in the testing set. Click on the choose button to find the naïve bayes classifier.
- (i) How much time did it take to build the classifier?
 - (ii) For the first example that was mis-classified, report the probability distribution over the classes. (Hint look under the <more options> button before you press start to print out the output predictions.)
 - (iii) What fraction of examples were correctly classified?
 - (iv) Report the confusion matrix on the testing set.
 - (v) Which class was mis-classified the most? How were the examples of that class usually classified?
- (b) Using the training set construct a neural network (multi-layer perceptron) and classify the examples in the testing set.
- (i) How much time did it take to build the classifier?
 - (ii) For the first example that was mis-classified, report the probability distribution over the classes.
 - (iii) What fraction of examples were correctly classified?
 - (iv) Report the confusion matrix on the testing set.
 - (v) Which class was mis-classified the most? How were the examples of that class usually classified?
- (c) Under the cluster tab, perform K-means clustering with 10 clusters and a seed of 10. (Hint: click on the “-n 2” part to modify the number of clusters.) Turn on Classes to clusters evaluation before pressing start.
- (i) Report the number of examples in each cluster.
 - (ii) Report the “confusion matrix” between clusters and classes.
 - (iii) What fraction of the examples were incorrectly clustered classes?
 - (iv) What class was incorrectly clustered the most? With which other classes was it most often clustered?