Name: _____

Problem	Points	Score
1(a)	5	
1(b)	5	
1(c)	5	
1(d)	5	
1(e)	10	
2(a)	10	
2(b)	10	
2(c)	10	
3(a)	10	
3(b)	10	
3(c)	10	
3(d)	10	
Total	100	

Notes:

(1) The exam is closed books and notes except for one double-sided sheet of notes.

(2) You are allowed to use a scientific calculator or the equivalent.

(3) Please indicate clearly your answer to the problem.

(4) Please try to make your solution legible and easy to follow. The better I can understand your thought process, the more generous I can be about partial credit. I will not give partial credit for ungrammatical sentences or fragmented answers. Please collect your thoughts and compose coherent answers.

(5) If you aren't sure how to work the details of a problem, at the very least write an outline of your solution indicating the step by step process that you think is needed to solve the problem.

Problem No. 1: A random variable X has the following distribution:

$$p(x) = \begin{Bmatrix} 1/2 & -1.0 \le x \le 1.0 \\ 0 & elsewhere \end{Bmatrix}.$$

A second random variable, Y, has the following distribution:

$$p(y) = \begin{Bmatrix} 1 & -1.0 \le y < -0.5 \\ 1 & 0.5 \le y < 1.0 \\ 0 & elsewhere \end{Bmatrix}.$$

(a) (5 pts) Compute the entropy of X and Y and compare. Explain why your answer makes sense.

(b) (5 pts) Assume $p(x, y)$ is given by:

$$p(x,y) = \begin{Bmatrix} 1/4 & -1.0 \le x < 1.0, -1.0 \le y < -0.5 \\ 1/4 & -1.0 \le x < 1.0, 0.5 \le y < 1.0 \\ 0 & elsewhere \end{Bmatrix}.$$

(c) (5 pts) Compute the joint entropy of X and Y.
(d) (5 pts) Compute the mutual information between X and Y.
(e) (10 pts) Explain why answers (a)-(d) make sense.

Problem No. 2: You are given an evaluation database of 10,000 images for an image recognition problem. Your baseline system was measured to have an error rate of 2.0%. Your new fancy machine learning algorithm delivers an error rate of X%.

(a) (10 pts) At a 95% confidence level, what must the value of X be for your results to be statistically significant?

(b) (10 pts) How would your answer to (a) change if we lowered the confidence level to 80%? Explain why your answer makes sense.

(c) (10 pts) Assume X is 1.9%. Assume a confidence level of 95%. How many images would you need in the evaluation set to make this result statistically significant?

Problem No. 3: In this problem, you will design several neural networks:

(a) (10 pts) Design a network that accepts binary data in the form of a 3-tuple (e.g., "101") and implements (or approximates as well as possible) the mapping shown to the right. Comment on the complexity of your network and how well it approximates this mapping. For the nonlinear unit at each node, use a hard limiter whose output is either 0 or 1.

Input	Output
000	0
001	1
010	1
011	0

(b) (10 pts) Compute the output of your network for previously unobserved patterns of "111", "110" and "100". Do these outputs make sense?

(c) (10 pts) Suppose you eliminate the third symbol in the input and disconnect whatever nodes you implemented to use that information. Suppose you apply "01", "10", "00" and "11" to this reduced network. Does it behave in a predictable manner? Explain (don't just tell me what the outputs are, explain why or why not the make sense). Note that you do not retrain the network.

(d) (10 pts) Suppose you replaced the hard limiter in the nonlinear unit with a sigmoid function. How would that change the system's ability to model this particular data?