

Name: _____

Problem	Points	Score
1	25	
2	25	
3	25	
4	25	
Total	100	

Notes:

- (1) The exam is closed books and notes except for one double-sided sheet of notes.
- (2) Please indicate clearly your answer to the problem.
- (3) If I can't read or follow your solution, it is wrong and no partial credit will be awarded.

(25 pts) Problem No. 1: A random variable, X , has a probability density functions defined as:

$$p(x) = \begin{cases} 1/4 & 0.0 \leq x \leq 0.5, \\ 7/4 & 0.5 \leq x \leq 1.0 \end{cases} .$$

A similar random variable, Y , has a probability density functions defined as:

$$p(y) = \begin{cases} 7/4 & 0.0 \leq y \leq 0.5, \\ 1/4 & 0.5 \leq y \leq 1.0 \end{cases} .$$

- Compute the entropy of X and Y . Explain why your answer makes sense.
- Assuming these variables are statistically independent, compute the joint entropy of X and Y .
- Assuming these variables are statistically independent, compute the mutual information between X and Y . Again, explain why this answer makes sense.

(25 pts) Problem No. 2: You are given an evaluation database of 1,000 images for an image recognition problem. Your baseline system was measured to have an error rate of 5.0%.

- Your proposed system, which is the centerpiece of your newest publication, delivers an error rate of 4.5%. At what level of confidence can you declare that the difference between the baseline system and your proposed system is statistically significant? Clearly explain any assumptions that are part of this calculation.
- How many images would you need in your evaluation set to declare that an error rate of 4.9% was better than the baseline system with a 95% confidence level (meaning this difference was statistically significant)?
- How would your answer to (b) change if we lowered the confidence level to 80%? Explain why your answer makes sense.

(25 pts) Problem No. 3: Given a set of measurements, $X = \{0, 1, 3, 4, 9, 10, 100, 101, 200, 201\}$:

- Cluster the data using a dendrogram approach (greedy hierarchical clustering). Clearly describe any assumptions you make to achieve your solution.
- Cluster the data using an agglomerative clustering approach.
- Compare your solutions. For example, if you decided to cluster the data into two clusters, how would these solutions differ?

(25 pts) Problem No. 4: In this problem, you will design several neural networks:

- 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

- Compute the output of your network for previously unobserved patterns of "111", "110" and "100". Do these outputs make sense?