

Name: _____

Problem	Points	Score
1	30	
2	30	
3	40	
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) Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit. Please carefully explain your solutions in well-written English.
- (4) If I can't read or follow your solution, it is wrong, and no partial credit will be awarded.
- (5) You are allowed to use MS Excel or the equivalent for calculations.

(30 pts) Problem No. 1: A discrete random variable, X , has a probability mass function (pmf):

$$p_k = \begin{cases} 1/16 & 0 \leq k < 4 \\ 3/16 & 4 \leq k < 8 \end{cases}.$$

A similar random variable, Y , has a probability mass function:

$$p_k = \begin{cases} 3/16 & 0 \leq k < 4 \\ 1/16 & 4 \leq k < 8 \end{cases}.$$

- Compute the entropy of X and Y . Explain why your answers 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.
- Suggest a shape for the pmf of X that would increase the mutual information. Justify your answer.

(30 pts) Problem No. 2: You are given an evaluation database of 2,000 images for an image recognition problem. Your baseline system was measured to have an error rate of 2.0%.

- Your proposed system, which is the centerpiece of your newest publication, delivers an error rate of 1.0%. 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 1.9% 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.

(40 pts) Problem No. 3: Consider a standard three-layer (multilayer) perceptron that uses an activation function given by:

$$y = \frac{1}{1 + \gamma + \alpha e^{-\beta \sum w_k x_k}}$$

- Derive an expression to update hidden layer weights assuming α , β and γ are constants.
- Improve the convergence properties of this system by adding regularization. Carefully explain all details of your approach and derive the necessary update equations.