

Name: _____

Problem	Points	Score
1(a)	10	
1(b)	10	
2(a)	10	
2(b)	10	
3(a)	10	
3(b)	10	
3(c)	10	
4(a)	10	
4(b)	10	
4(c)	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: Prove the following:

- (a) (10 pts) $H(X, Y) = H(X) + H(Y|X)$: the joint entropy in terms of the conditional entropy
- (b) (10 pts) $I(X; X) = H(X)$: mutual information between a random variable X and itself

Problem No. 2: Mary runs an experiment using an evaluation set of 100 samples and declares her new algorithm is better than the baseline because the baseline error rate was 1.00% and her new algorithm yielded an error rate of 0.95%. She says she is 95% confident of this.

- (a) (10 pts) Do you believe her? Be as specific as possible. (Note that you do not need to do detailed calculations, just support your argument with equations.)
- (b) (10 pts) Suppose you disagree with her. Estimate the increase in sample size needed to allow her to make this claim. Again, you don't need to do detailed calculations, but you need to support your estimate with equations and explain how you arrived at it.

Problem No. 3: You are given two data sets:

Class 1: $\{[1, 1], [2, 1], [1, 2], [2, 2]\}$
 Class 2: $\{[-1, -1], [-2, -1], [-1, -2], [-2, -2]\}$

- (a) (10 pts) Cluster each class using a top-down binary tree (Linde-Buzo-Gray) algorithm. Provide a dendrogram that demonstrates the resulting clustering of each class.
- (b) (10 pts) Using two clusters per class, classify the points $[0.5, 0.5]$ and $[-0.5, -0.5]$ using a Euclidean distance metric. Indicate whether or not they are correctly classified. Would this result change if we used a Mahalanobis distance measure?
- (c) (10 pts) Estimate the error rate for an evaluation set that consisted of data uniformly distributed on the rectangular region bounded by: $\{[0.5, 0.5], [0.5, -0.5], [-0.5, -0.5], [-0.5, 0.5]\}$.

Problem No. 4: In this problem, you will compare two classification approaches:

- (a) (10 pts) Design a network that classifies the data in the table to the right.
- (b) (10 pts) Design a system based on K-MEANS clustering that classifies this data. Use a "Hamming Distance" as your distance metric (e.g., count the number of 1's and 0's that are different).
- (c) (10 pts) Compare and contrast these networks. Discuss the pros and cons. Don't simply describe what they do. Analyze why they are different.

Set	Class	Value
Train	0	0000
Train	0	0001
Train	0	0010
Train	0	0011
Train	1	1111
Train	1	1110
Train	1	1101
Train	1	1100
Eval	0	1010
Eval	1	0101