

Name: _____

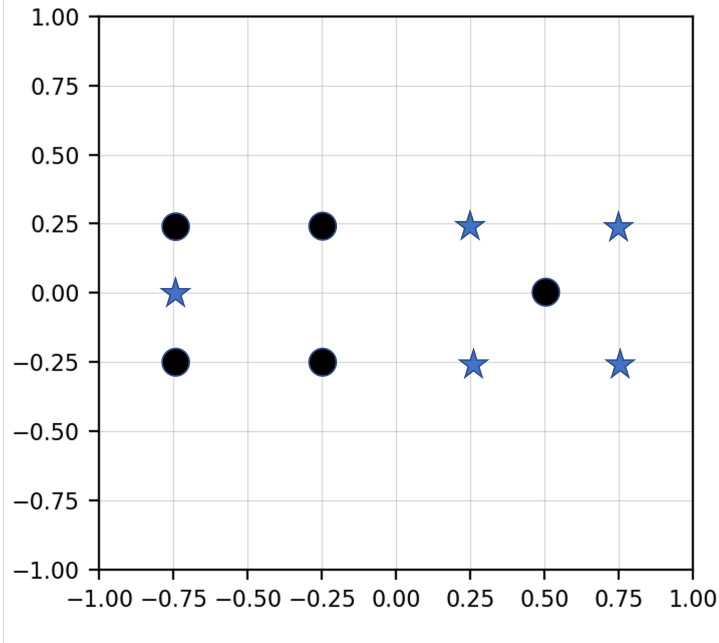
Problem	Points	Score
1(a)	10	
1(b)	10	
1(c)	10	
2	20	
3(a)	10	
3(b)	10	
4(a)	10	
4(b)	10	
4(c)	10	
Total	100	

Notes:

- (1) The exam is closed books and notes. You are allowed on double-sides 8.5x11 in. page of notes. If you don't have an equation that you think you need, please ask me during the exam.
- (2) Please clearly indicate your answer to the problem.
- (3) Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit.
- (4) If I can't read or follow your solution, it is wrong, and no partial credit will be awarded.

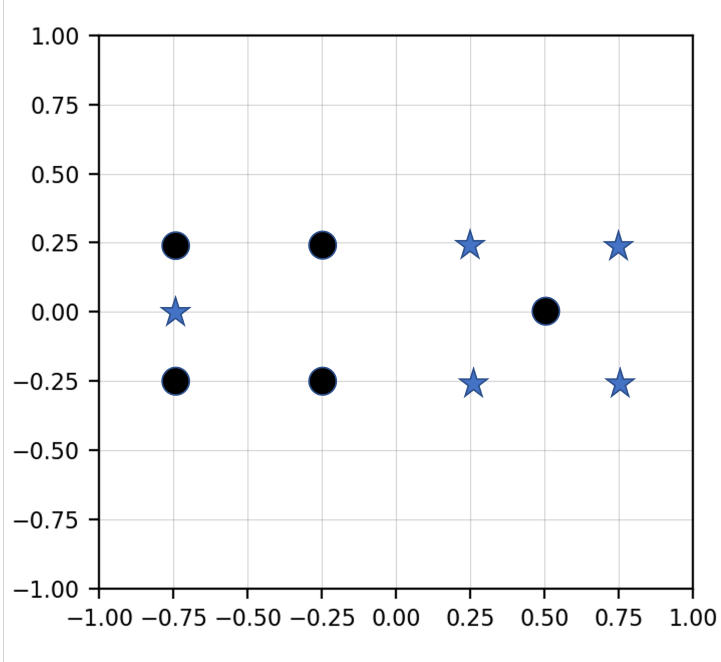
Problem No. 1: Given the data shown below:

- (a) (10 pts) Draw the decision surface you would obtain if you applied the k-nearest neighbor algorithm (KNN) with $K = 1$ to this data. Justify your result with a detailed explanation.



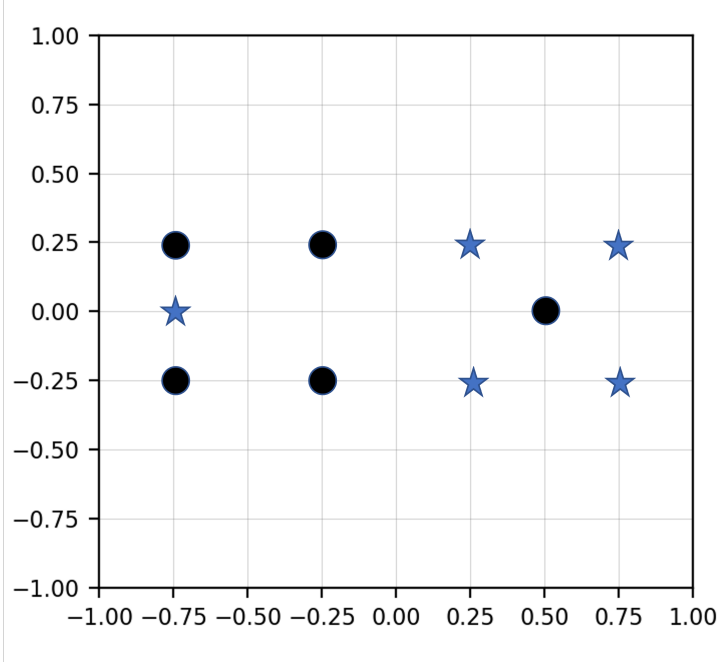
Justification:

- (b) (10 pts) Draw the decision surface you would obtain if you applied the k-nearest neighbor algorithm (KNN) with $K = 3$ to this data. Use a majority vote rule. Justify your result with a detailed explanation.



Justification:

(c) (10 pts) What would Quadratic Discriminant Analysis (QDA) do in this case? Sketch the shape of the decision surface and explain any underlying calculations (e.g. what does the support region for the underlying multivariate Gaussian distribution look like?).



Justification:

(20 pts) Problem No. 2: Derive a gradient descent algorithm to optimize the weights for the following function:

$$f(x) = w_0 + w_1 x_1 + w_2 x_2^2$$

Explain your approach fully including any assumptions you make about the loss function, etc.

Problem No. 3: You are given the data shown to the right as a training set for two-class classification problem.

(a) (10 pts) Using the information theory principles discussed in class, and using quantitative measures based on concepts such as entropy, determine which feature, F1 or F2, is the more reliable feature.

Index	Class	F1	F2
1	1	T	T
2	1	T	F
3	0	F	F
4	0	F	T
5	0	F	T

(b) (10 pts) You are given two additional measurements shown in the second table to the right. Again, using information theory principles, quantify the impact this data has on your ability to build a classification system. Do these measurements reinforce what you learned from the first table? Do they contradict what you have learned?

Index	Class	F1	F2
6	1	F	T
7	0	T	F

Your answers will be evaluated not based on casual observations (which one might argue are obvious) but based on your ability to justify these observations using calculations based on information theory principles.

Problem No. 4: Using the equation we discussed in class for determining statistical significance:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{N} + \frac{\hat{p}_2(1-\hat{p}_2)}{N}\right)}}$$

(a) (10 pts) Determine the number of independent samples required to demonstrate that your new research system, which gives you an error rate of 9.5%, is better than the baseline system, which has an error rate of 10.0%. Use a 95% confidence value (a 95% confidence value corresponds to a range of [-1.96, 1.96] on a Normal distribution).

(b) (10 pts) Would the margin of error change if the baseline system had an error rate of 20%, and your system had an error rate of 19.5%? If so, explain why. (Don't simply tell me what the calculations show. Explain what about the theory behind these calculations determines any changes in the margin of error.)

(c) (10 pts) Suppose you observe that the probability distribution for your observed error rates is not Gaussian. How might this influence your confidence intervals? Hint: Think about the principle of maximum entropy.