

Name: _____

Problem	Points	Score
1	30	
2(a)	10	
2(b)	10	
2(c)	10	
3(a)	10	
3(b)	10	
3(c)	20	
Total	100	

Notes:

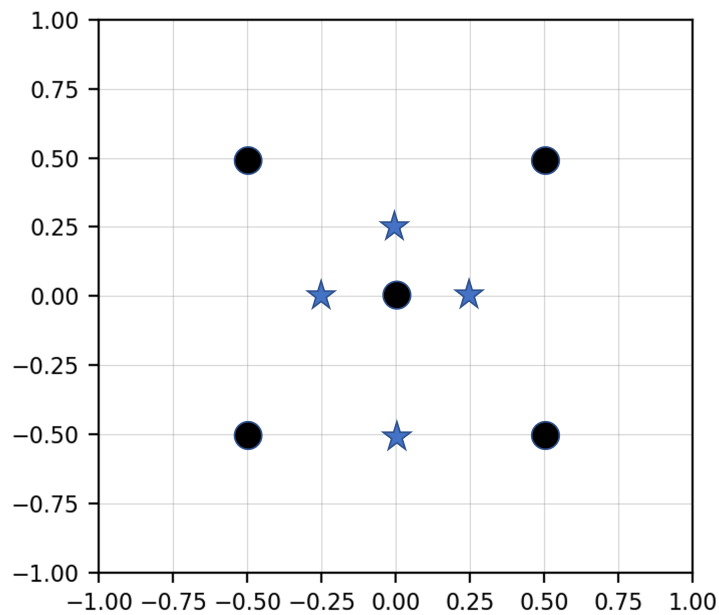
- (1) The exam is closed books and notes. You are allowed on double-sides 8.5x11 in. page of notes. If you don't have an equation that you think you need, please ask me during the exam.
- (2) Please clearly indicate your answer to the problem.
- (3) Note that ungrammatical sentences, incoherent statements, 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 (30 points): Devise an algorithm that uses dynamic programming to solve the spelling correction problem. Given two words: “jane” and “jnae”, compute the distance between them and/or the optimal alignment. You must allow for transpositions of one character (but multiple transpositions can occur in a single string – see below). Demonstrate how your algorithm works by setting up a grid and showing how your algorithm proceeds through the grid to find the optimal solution. Be sure to carefully describe your cost functions, node costs, etc.

Can you solve a second case - “spring” vs. “spirgn” – using the same algorithm? What about this one – “aaaabcbdddefgggg” vs. “aaaacbddddfeegg”?

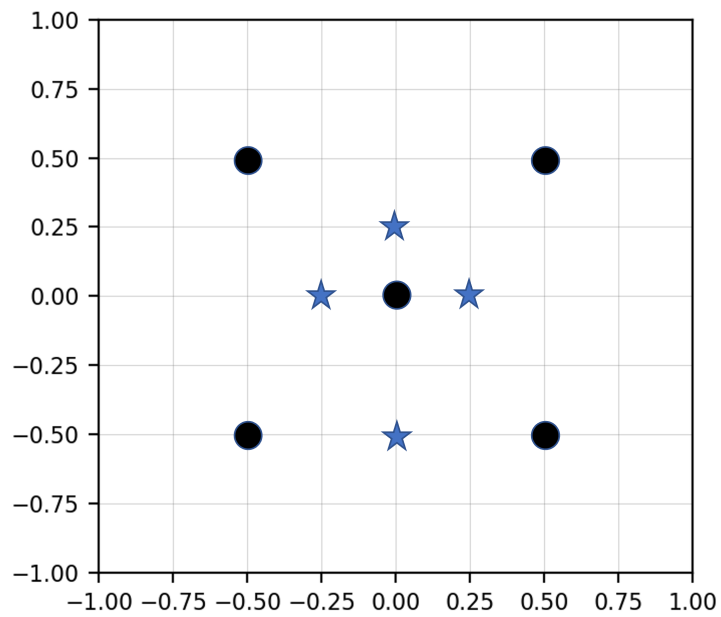
Problem No. 2 (30 points): For the examples shown, draw the decision surfaces each algorithm would produce, and explain how you arrived at this solution:

(a) (10 pts) Principal Component Analysis (aka, class-independent/pooled covariance):



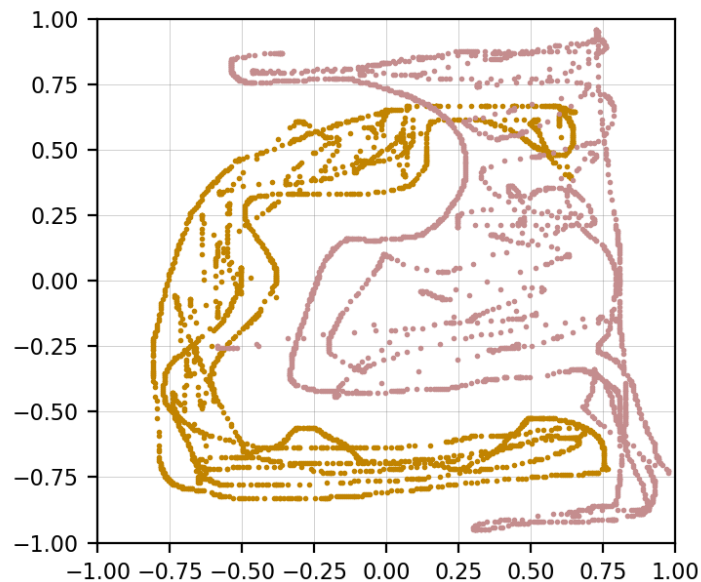
Justification:

(b) (10 pts) K-MEANS with two clusters ($K = 2$) per class:

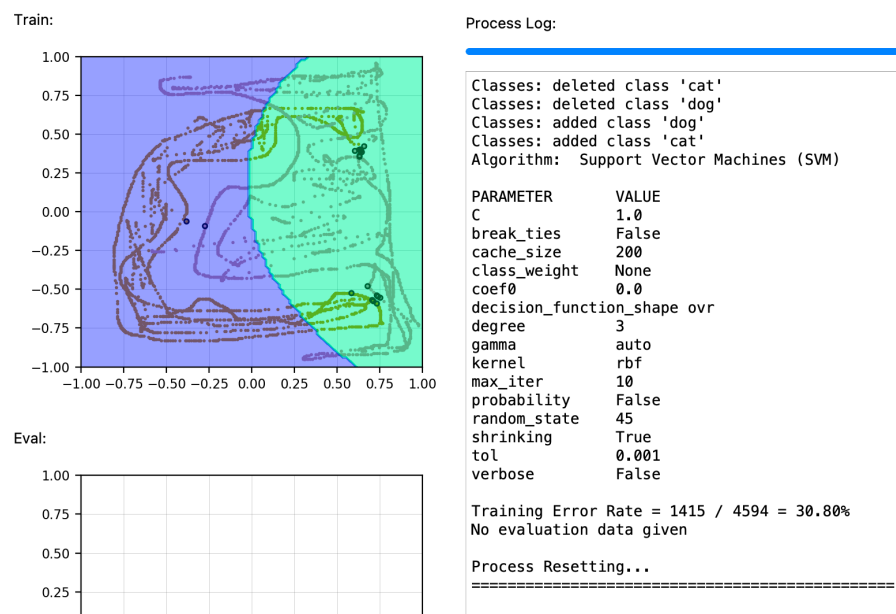


Justification:

(c) (10 pts) Here is a very messy data set:



My fantastic research students delivered to me a version of Support Vector Machines that they claim is fully debugged. Below is what I got on closed-set testing:



Is this correct? Are there different parameter settings or variants of the algorithm that might do better?

Problem No. 3 (40 pts): A famous paper you are reading claims a fundamental algorithm improvement to a system based on a standard neural network, reducing state of the art from an error rate of 1.0% to 0.8% – a 20% relative reduction in error rate. The sample size is 10,000 data points.

(a) (10 pts) Is this result statistically significant? Explain in as great a deal as possible. You might find

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

(b) (10 pts) Is (a) completely specified, or is other information needed to rigorously answer the question?

(c) (20 pts) Another set of authors comes along and claims an astounding breakthrough – 0.1% error. To achieve this result, the new authors introduced a new feature to the system previously described by the original authors. The training and evaluation databases were augmented by this feature. The machine learning system introduced by the new authors (which was based on quantum computing ☺) had previously been shown to not be significantly better on 10 other large, publicly available databases. But on this specific dataset, with this new feature vector, the new set of authors produced astounding results.

Using the information theory principles discussed in class, explain some of the things you need to consider in evaluating whether this new system was truly a breakthrough. Is it possible this result was fraudulent? Why? How would you go about proving that?