CalPolyPomona

CS 4210 – Assignment #3
Maximum Points: 100 pts.

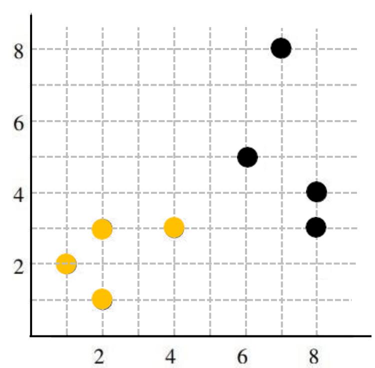
Bronco ID: |__|__|__|__|__|__|__|__|__|

Last Name: ________________________

First Name: ________________________

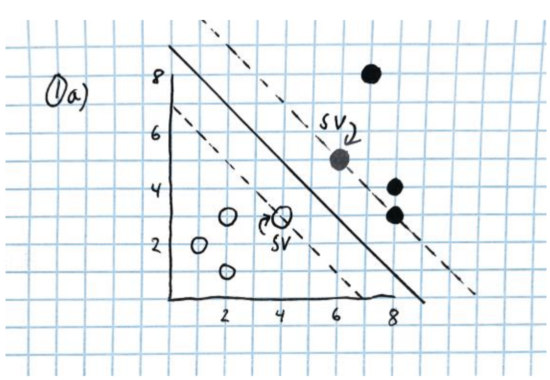
Note 1: Your submission header must have the format as shown in the above-enclosed rounded rectangle.
Note 2: Homework is to be done individually. You may discuss the homework problems with your fellow students, but you are NOT allowed to copy – either in part or in whole – anyone else's answers.
Note 3: Your deliverable should be a .pdf file submitted through Gradescope until the deadline. Do not forget to assign a page to each of your answers when making a submission. In addition, source code (.py files) should be added to an online repository (e.g., github) to be downloaded and executed later.
Note 4: All submitted materials must be legible. Figures/diagrams must have good quality.
Note 5: Please use and check the Canvas discussion for further instructions, questions, answers, and hints. The bold words/sentences provide information for a complete or accurate answer.

1. [20 points] Say you are given the training dataset shown below. This is a binary classification task in which the instances are described by two integer-valued attributes.



Dataset

a. [2 points] Draw the decision boundary and its parallel hyperplanes for a linear SVM with maximum margin (hard margin formulation) and identify the support vectors.



b. [2 points] If a black circle is added as a training sample in the position (7,5), does this affect the previously learned decision boundary? Explain why.

This would not affect the decision boundary because it would not be any closer to the opposite yellow support vector at (4, 3) than the current black support vector at (6, 5).

c. [2 points] If a yellow circle is added as a training sample in the position (4,2), does this affect the previously learned decision boundary? Explain why.

This would also not affect the decision boundary because it would not be any closer to the opposite black support vector at (6, 5) than the current yellow support vector at (4, 3).

d. [2 points] If a black circle is added as a test sample in the position (7,5), will this sample be classified correctly according to the previously learned decision boundary? Explain why.

This sample would be classified correctly because it is outside the margin of the separating hyperplane on the side of the black support vector.

e. [2 points] If a black circle is added as a test sample in the position (6,4), will this sample be classified correctly according to the previously learned decision boundary? Explain why.

This will be classified correctly. Although it is inside the margin of the separating hyperplane, it is still on the side of the decision boundary with the black support vector. Therefore, the new point will be correctly classified as black.

f. [2 points] If a yellow circle is added as a test sample in the position (4,2), will this sample be classified correctly according to the previously learned decision boundary? Explain why.

This sample will be classified correctly. It is outside the margin of the separating hyperplane on the side of the yellow support vector.

g. [2 points] If a yellow circle is added as a test sample in the position (5,3), will this sample be classified correctly according to the previously learned decision boundary? Explain why.

This will be classified correctly. Although it is inside the margin of the separating hyperplane, it is still on the side of the decision boundary with the yellow support vector. Therefore, the new point will be correctly classified as yellow.

h. [2 points] If a black circle is added as a test sample in the position (5,3), will this sample be classified correctly according to the previously learned decision boundary? Explain why.

This sample will not be classified correctly. The point would be on the bottom side of the decision boundary, and be classified based on that side's support vector, which is yellow, not black.

i. [2 points] If a yellow circle is added as a test sample in the position (6,4), will this sample be classified correctly according to the previously learned decision boundary? Explain why.

This sample will not be classified correctly. The point would be on the top side of the decision boundary, and be classified based on that side's support vector, which is black, not yellow.

j. [2 points] If a black circle is added as a training sample in the position (4,4), how this will affect the decision boundary if C = 1 and C = ∞? Consider the soft margin formulation.

With the C=1 in a soft margin formulation, the new point would not affect the decision boundary because the error function would allow it within the margin of separating hyperplanes. However if C=∞, the margin would absolutely change. The black circle at (4, 4) would become a new support vector and change the decision boundary.

2. [20 points] Consider the following 1-dimensional data with two classes:

x	-3	0	1	2	3	4	5
Class	$-$	$-$	$+$	$+$	$+$	$+$	$+$

a. [5 points] Find the decision boundary of a linear SVM on this data (hard-margin formulation) and identify the support vectors (write the x coordinate to provide your answer).

Decision Boundary = $\frac{1}{2}$

Support Vectors: $x = 0, x = 1$

b. [5 points] Find the solution parameters w and b for this linear SVM and the width of the margin. Hint: place the identified support vectors (positive and negative) into the formula $y_i(w.x_i + b) = 1$ since you know this formula holds for them.

$y_i(w \cdot x_i + b) = 1$
$(-1)(w \cdot (0) + b) = 1$
$b = -1$

$(1)(w \cdot (1) + (-1)) = 1$
$w = 2$

Lets see if this holds for

$w \cdot x + b = 0$
$2 \cdot \left(\frac{1}{2}\right) - 1 = 0$

$0 = 0$
True!

c. [5 points] Show mathematically that the SVM classifications for the test data {-1.5, 1.5} are negative and positive respectively.

$$y = \begin{cases} 1, & w \cdot z + b > 0 \\ -1, & w \cdot z + b > 0 \end{cases}$$

$$y = \begin{cases} 1, & 2 \cdot z - 1 > 0 \\ -1, & 2 \cdot z - 1 > 0 \end{cases}$$

$$y(-1.5) = 2 \cdot (-1.5) - 1 = -4 = -1$$
$$y(1.5) = 2 \cdot (1.5) - 1 = 2 = 1$$

d. [5 points] Suppose we remove the point (1,+) from this training set and train the SVM again. Find the new values of the solution parameters *w* and *b* and the width of the margin.

$y_i(w \cdot x_i + b) = 1$

$(-1)(w \cdot (0) + b) = 1$
$b = -1$

$(1)(w \cdot (2) + (-1)) = 1$
$w = 1$

$margin\ distance = d = \frac{2}{||w||}$

$d = \frac{2}{1} = 2$

3. [20 points] The quadratic kernel $K(x, y) = (x.y + 1)^2$ should be equivalent to mapping each x into a six-dimensional space where

$$\Phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$$

for the case where $x = (x_1, x_2)$. Demonstrate this equivalence by answering the following questions while using the data points: $A = (1,2)$, $B = (2,4)$.

a. [5 points] $\Phi(A)$

$$\Phi(\mathrm{A}) = (1,4,2\sqrt{2}, \sqrt{2}, 2\sqrt{2}, 1)$$

b. [5 points] $\Phi(B)$

$$\Phi(\mathrm{B}) = (4,16,8\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}, 1)$$

c. [5 points] $\Phi(A)\cdot\Phi(B)$

$$\Phi(\mathrm{A}) \cdot \Phi(\mathrm{B}) = (1 * 4) + (4 * 16) + \left(2\sqrt{2} * 8\sqrt{2}\right) + \left(2\sqrt{2} * \sqrt{2}\right) + \left(2\sqrt{2} * 4\sqrt{2}\right) + 1$$
$$\Phi(\mathrm{A}) \cdot \Phi(\mathrm{B}) = 121$$

d. [5 points] $K(A, B)$. Hint: your answers for (c) and (d) should be the same. By using the kernel function, SVM "cheats" and performs significantly fewer calculations (kernel trick).

$$K(A, B) = \left((1 * 2) + (2 * 4) + 1\right)^2$$
$$K(A, B) = 121$$

4. [15 points] Complete the Python program (svm.py) that will read the file optdigits.tra (3,823 samples) that includes training instances of handwritten digits (optically recognized). Read the file optdigits.names to get detailed information about this dataset. Also, check the file optdigits-orig.tra and optdigits-orig.names to see the original format of this data, and how it was transformed to speed-up the learning process (pre-processing phase). Your goal is to build multiple SVM classifiers using this data. You will simulate a grid search, trying to find which combination of four SVM hyperparameters (c, degree, kernel, and decision_function_shape) leads you to the best prediction performance. To test the accuracy of those distinct models, you will use the file optdigits.tes (1,797 samples).

The most accurate model was with the combination :

C = 10
Degree = 1
Kernel = rbf
Decision Function Shape = ovo

5. [25 points] Consider the dataset below.

Outlook	Temperature	PlayTennis
Sunny	Hot	No
Overcast	Cool	Yes
Overcast	Hot	Yes
Rain	Cool	No
Overcast	Mild	Yes

We will apply *steady state* Genetic Algorithms ($r = 0.5$) to solve this classification problem. If the instance matches the rule pre-condition of the chromosome, you predict according to its post-condition, otherwise you predict the opposite class defined by the chromosome (there must be a prediction for each instance then). Follow the planning below to build your solution.

- Representation: single if-then rule by using bit strings (binary encoding).
    - Outlook $<Sunny, Overcast, Rain>$
    - Temperature $<Hot, Mild, Cool>$
    - Examples:
    - $\text{Outlook} = Overcast \rightarrow 010$
    - $\text{Outlook} = Overcast \vee \text{Rain} \rightarrow 011$
    - $\text{Outlook} = Sunny \vee Overcast \vee \text{Rain} \rightarrow 111$
    - $\text{Outlook} = (Overcast \vee Rain) \wedge (\text{Temperature} = Hot) \rightarrow 011100$
    - $\text{Outlook} = Sunny \wedge \text{Temperature} = Hot$ then $\text{PlayTennis} = yes \rightarrow 1001001$
- Initial population (Chromosomes) : (C_1=100|100|1, C_2=010|010|1, C_3=101|100|0, C_4=110|110|0). Population size should remain the same (4 individuals) over time.
- Fitness function: accuracy
- Penalty criterion: no penalty.
- Selection method: The best two chromosomes are carried over to the next generation. The other two are selected for crossover by using the roulette wheel (simulation).
- Crossover strategy: single-point crossover with mask 1110000 in the 1st generation, two-point crossover with mask 0001100 in the 2nd generation. Use the following chromosomes to perform crossover (simulating the process of spinning the roulette wheel) according to the relative fitness (sectors of a roulette wheel), generating two offspring for each crossover:
    - (1st and 3rd) in the 1st generation
    - (1st and 2nd) in the 2nd generation
- Mutation: on the 6th bit of the chromosome(s) 1011000 selected/produced during the 2nd generation.
- Termination criteria: accuracy = 1.0. **Return the corresponding chromosome(s) – this will be your model.**

Solution format:

Fitness(C_1) = 1/5
Fitness(C_2) = 3/5
Fitness(C_3) = 4/5
Fitness(C_4) = 2/5

1st generation (C_2, C_3, C_5, C_6):

Pr(C_1) = 0.1 (4th)
Pr(C_4) = 0.2 (3rd)
Pr(C_2) = 0.3 (2nd)
Pr(C_3) = 0.4 (1st)

C_1 = 1001001 → C_5 = 1001000
C_3 = 1011000→ C_6 = 1011001

Fitness(C_2) = 3/5
Fitness(C_3) = 4/5
Fitness(C_5) = 4/5
Fitness(C_6) = 1/5

2nd generation (C_3, C_5, C_7, C_8)

Pr(C_6) = 0.083 (4th)
Pr(C_2) = 0.25 (3rd)
Pr(C_3) = 0.33 (2nd)
Pr(C_5) = 0.33 (1st)

C_3 = 1011000→ C_7 = 1011000
C_5 = 1001000→ C_8 = 1001000

Applying mutation on 1011000

Mutated C_7 =1011010

Fitness(C_3) = 4/5
Fitness(C_5) = 4/5
Fitness(C_7) = 5/5
Fitness(C_8) = 4/5

Final answer: C_7

(1011010)

Important Note: Answers to all questions should be written clearly, concisely, and unmistakably delineated. You may resubmit multiple times until the deadline (the last submission will be considered).

NO LATE ASSIGNMENTS WILL BE ACCEPTED. ALWAYS SUBMIT WHATEVER YOU HAVE COMPLETED FOR PARTIAL CREDIT BEFORE THE DEADLINE!