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. [15 points] Considering the 1-D dataset below and the following bootstrap samples (bagging rounds 1 to 5) randomly generated during the bagging process. Show how a bagging algorithm can perfectly classify this data by **drawing** and **writing** the decision stumps for each round, the summary table of the trained decision stumps, and the combination table of your base classifiers with the final predictions. Hint: you might need to test alternative but equally accurate decision stumps on your training set to get maximum accuracy.

x	1	2	3	4	5
y	-1	-1	1	1	-1

Dataset

x	1	1	2	4	5

Round 1

x	3	3	4	4	5

Round 2

x	1	2	2	5	5

Round 3

x	1	3	4	4	5

Round 4

x	1	2	3	3	4

Round 5

CS 4210 - Assignment 5

1.

x	1	1	2	4	5
y	-1	-1	-1	1	-1

Round 1

Split point: $x = 5.5$

$x < 5.5, -1 \quad x \geq 5.5, 1$

x	3	3	4	4	5
y	1	1	1	1	-1

Round 2

Split point: $x = 4.5$

$x < 4.5, 1 \quad x \geq 4.5 -1$

x	1	2	2	5	5
y	-1	-1	-1	-1	-1

Round 3

Split point: $x = .5$

$x < .5, -1 \quad x \geq .5 -1$

x	1	3	4	4	5
y	-1	1	1	1	-1

Round 4

Split point: $x = 2$

$x < 2, -1 \quad x \geq 2 \; 1$

x	1	2	3	3	4
y	-1	-1	1	1	1

Round 5

Split point: $x = 2.5$

$x < 2.5, -1 \quad x \geq 2.5, 1$

Round	x=1	x = 2	x = 3	x = 4	x = 5
1	-1	-1	-1	-1	-1
2	1	1	1	1	-1
3	-1	-1	-1	-1	-1
4	-1	1	1	1	1
5	-1	-1	1	1	1
Res	-1	-1	1	1	-1

2. [12 points] Considering the different 1-D dataset below and the following rounds from 1 to 3 randomly generated during the boosting process. Show how a boosting algorithm can perfectly classify this data by **drawing** and **writing** the decision stumps and weights for each round, the summary table of the trained decision stumps, and the combination table of your base classifiers with the weighted final predictions. Hint: there is a single best decision stump (more accurate) for each round.

x	1	2	3	4	5	Dataset
y	1	1	-1	-1	1	

x	1	2	3	4	4	Round 1

x	5	5	5	5	5	Round 2

x	3	3	4	4	5	Round 3

2.

x	1	2	3	4	4
y	1	1	-1	-1	-1

Round 1

$x = 2.5$

$x < 2.5,\ 1 \quad x \geq 2.5\ -1$

$$\epsilon_1 = \frac{1}{5}\Big(\underbrace{.2 \cdot 1}_{5 \text{ misclassified}}\Big) = .04$$

$$\alpha_1 = \frac{1}{2}\ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) = \frac{1}{2}\ln\left(\frac{1-.04}{.04}\right) = 1.59$$

$$w_1^1 = w_2^1 = w_3^1 = w_4^1 = \frac{.2}{Z_t} \cdot e^{-1.59} = \frac{.0408}{Z_t}$$

$$w_5^1 = \frac{.2 \cdot e^{1.59}}{Z_t} = \frac{.981}{Z_t}$$

$$\underbrace{\frac{4(.0408)}{Z_t} + \frac{.981}{Z_t}}_{\text{since } \sum w^{(t)} = 1} = 1 \qquad Z_t = 1.1442$$

$$w_1^1 = w_2^1 = w_3^1 = w_4^1 = \frac{.0408}{1.1442} = .0357$$

$$w_5^1 = \frac{.981}{1.1442} = .857$$

x	5	5	5	5	5
y	1	1	1	1	1

$x = 5.5$

$x < 5.5,\ 1 \quad x \geq 5.5\ 1$

$$\epsilon_2 = \frac{1}{5}\Big(\underbrace{.0357 \cdot 1 + .0357 \cdot 1}_{x=3,4 \text{ are misclassified}}\Big) = .01428$$

$$\alpha_2 = \frac{1}{2}\ln\left(\frac{1-.01428}{.01428}\right) = 2.117$$

$$w_1^2 = w_2^2 = \frac{.0357}{Z_t} \cdot e^{-2.117} \qquad w_3^2 = w_4^2 = \frac{.0357}{Z_t} e^{2.117} \qquad w_5^2 = \frac{.857}{Z_t} e^{-2.117}$$

$w_1^2 = w_2^2 = \frac{4.30 \times 10^{-3}}{Z_t}$ $w_3^2 = w_4^2 = \frac{.297}{Z_t}$ $w_5^2 = \frac{.103}{Z_t}$

$Z_t = 2(4.30 \times 10^{-3}) + 2(.297) + .103 = .7056$

$w_1^2 = w_2^2 = 6.094 \times 10^{-3}$ $w_3^2 = w_4^2 = .4209$ $w_5^2 = .146$

x	3	3	4	4	5
y	-1	-1	-1	-1	1

Round 3

$x = 4.5$

$x < 4.5$ -1 $x \geq 4.5$ 1

$\epsilon_3 = \frac{1}{5}\Big(\underbrace{6.094 \times 10^{-3} \cdot 1 + 6.094 \times 10^{-3} \cdot 1}_{x=1,2 \text{ are misclassified}}\Big) = 2.4376 \times 10^{-3}$

$\alpha_3 = \frac{1}{2} \ln\left(\frac{1 - 2.4376 \times 10^{-3}}{2.4376 \times 10^{-3}}\right) = 3$

$w_1^3 = w_2^3 = \frac{6.094 \times 10^{-3}\, e^{3}}{Z_t} = \frac{.122}{Z_t}$

$w_3^3 = w_4^3 = \frac{.4209}{Z_t} e^{-3} = \frac{.021}{Z_t}$ $w_5^3 = \frac{.146}{Z_t} e^{-3} = \frac{7.27 \times 10^{-3}}{Z_t}$

$Z_t = 2(.122) + 2(.021) + 7.27 \times 10^{-3}$

$= .293$

$w_1^3 = w_2^3 = .416$ $w_3^3 = w_4^3 = .0717$ $w_5^3 = .0248$

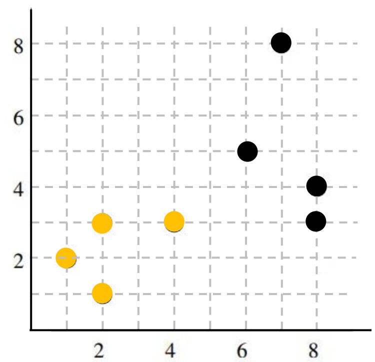
Round	Split point	Left class	Right class	Alpha
1	2.5	1	-1	1.59
2	5.5	1	1	2.117
3	4.5	-1	1	3

Round	x=1	x=2	x=3	x=4	x=5
1	1	1	-1	-1	-1
2	1	1	1	1	1
3	-1	-1	-1	-1	1
Pred	1	1	-1	-1	1

3. [15 points] Complete the Python program (bagging_random_forest.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 a base classifier by using a single decision tree, an ensemble classifier that combines multiple decision trees, and a Random Forest classifier to recognize those digits. To test the accuracy of those distinct models, you will use the file optdigits.tes (1,797 samples).

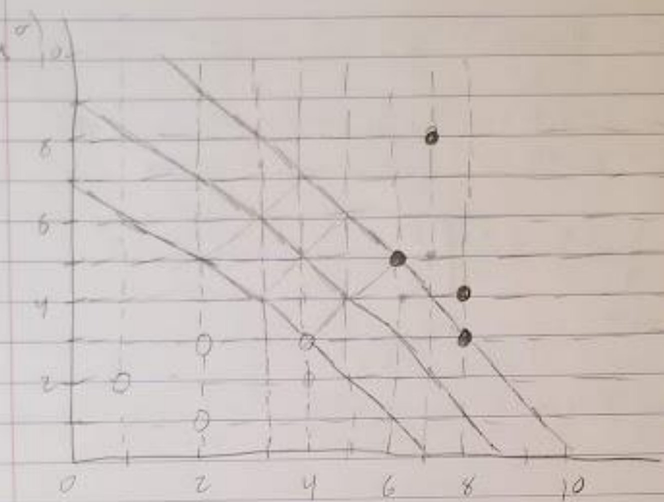
https://github.com/SeveralCube22/CS4210_Assignments/tree/master/Assignment%203

4. [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.



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.
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.
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.
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.
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.
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.
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.
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.
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.

a)



Instances @ (4,3) (6,5) are support vectors

b) No, this will violate the constraint that all values of one class should be on or above margin as (6,5) will be within the margin

c) No, this will violate the constraint that all values of one class should be on or below the margin as (3,3) will be within the margin

d) Yes this point is above the margin indicating its class is black.

e) Yes, this point is above the decision boundary so it's class is equal to black

f) Yes, this point is below the margin indicating its class is yellow.

g) Yes, this point is below the decision boundary so it's predicted to be yellow.

h) No, this point is below the decision boundary, so it is predicted to be yellow.

i) No, this point is above the decision boundary so it is predicted to be black

j) C=1, Margin would be large and (4,4) would be misclassified. C=10, Margin will be small with (4,4) on margin.

5. [11 points] Consider the following 1-dimensional data with two classes:

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

a. [3 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).
b. [3 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.
c. [2 points] Show mathematically that the SVM classifications for the test data {-1.5, 1.5} are negative and positive respectively.
d. [3 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.

5. a) support vectors: $x = 0, 1$

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

$w(1) + b = 1$

$w - 1 = 1$

$w = 2$

width $= \frac{2}{\|w\|} = \frac{2}{2} = 1$

c) $2(-1.5) - 1 = -4 \quad -4 \leq -1$ so negative

$2(1.5) - 1 = 2 \quad 2 \geq 1$ so positive

d) Support vectors: $x = 0, 2$

$w(0) + b = -1$

$b = -1$

$w(2) - 1 = 1$

$2w = 2$

$w = 1$

width $= \frac{2}{\|w\|} = \frac{2}{1} = 2$

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

$K(x, y) = (x \cdot y + 1)^2$

$A = (1, 2) \quad B = (2, 4)$

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

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

c) $\Phi(A) \cdot \Phi(B) = 4 + 64 + 32 + 4 + 16 + 1 = 121$

d) $K(A, B) = \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 1 \right)^2$

$= ((2 + 8) + 1)^2 = 11^2 = 121$

6. [12 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. [3 points] $\Phi(A)$
   b. [3 points] $\Phi(B)$
   c. [3 points] $\Phi(A)\Phi(B)$
   d. [3 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).

7. [15 points] Complete the Python program (svm.py) that will also read the file optdigits.tra to build multiple SVM classifiers. 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 also use the file optdigits.tes. You should update and print the accuracy, together with the hyperparameters, when it is getting higher.

https://github.com/SeveralCube22/CS4210_Assignments/tree/master/Assignment%203

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!