Bronco ID: 013001924
Last Name: Nalbandian
First Name: Andre

Assignment 3

Wednesday, October 13, 2021 10:27 PM

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	Dataset
y	-1	-1	1	1	-1	

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

Dataset

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

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

$x \le 3 \to y = -1$
$x > 3 \to y = 1$

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

$x \le 4.5 \to y = 1$
$x > 4.5 \to y = -1$

x	1	2	2	5	5	Round 3
y	-1	-1	-1	-1	-1	

$x \le .5 \to y = -1$
$x > .5 \to y = -1$

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

$x \le 4.5 \to y = 1$
$x > 4.5 \to y = -1$

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

$x \le 2.5 \to y = -1$
$x > 2.5 \to y = 1$

Summary

Round	Split	Left	Right
1	3	-1	1
2	4.5	1	-1
3	.5	-1	-1
4	4.5	1	-1
5	2.5	-1	1

R	1	2	3	4	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
Sum	-1	-1	1	3	-1
Sign	-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.

Dataset

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

Round 1

x	1	2	3	4	4

Round 2

x	5	5	5	5	5

Round 3

x	3	3	4	4	5

Dataset

Dataset

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

[redacted]

Round 1

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

$x \le 2.5 \to y = 1$
$x > 2.5 \to y = -1$

[redacted]

Round 2

x	5	5	5	5	5
	1	1	1	1	1

$x < 4.5 \to y = 1$
$x \ge 4.5 \to y = 1$

[redacted]

Round 3

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

$x \le 4.5 \to y = -1$
$x > 4.5 \to y = 1$

Weights

R	1	2	3	4	5
1	.2	.2	.2	.2	.2
2	.036	.036	.036	.036	.853
3	.006	.006	.424	.424	.141

	Split	Left	Right	Alpha
1	2.5	1	-1	1.59
2	4.5	1	1	2.12
3	4.5	-1	1	3.01

	1	2	3	4	5
1	1	1	-1	-1	-1
2	1	1	1	1	1
3	-1	-1	-1	-1	1
Sum	.7	.7	-2.48	-2.48	3.54
Sign	1	1	-1	-1	1

$\varepsilon_1 = \frac{1}{5}[.2 * 1] = .04$

$\alpha_1 = \frac{1}{2}\ln[(1 - .04)/.04] = 1.59$

$W_1^2 = W_2^2 = W_3^2 = W_4^2 = \frac{.2 \times e^{-1.59}}{Z_1} = \frac{.041}{1.15} = .036$

$W_5^2 = \frac{.2 * e^{1.59}}{Z_1} = \frac{.981}{1.15} \approx .853$

$Z_1 = .041 * 4 + .981 = 1.15$

$\varepsilon_2 = \frac{1}{5}[.036 * 1 + .036 * 1] = .014$

$\alpha_2 = \frac{1}{2}\ln((1 - .014)/(.014)) = 2.12$

$W_1 = W_2 = \frac{.036 \times e^{-2.12}}{Z_2} = \frac{.004}{Z_2} \approx .006$

$W_5 = \frac{.853 \times e^{-2.12}}{Z_2} = \frac{.1}{Z_2} = .141$

$W_3 = W_4 = \frac{.036 \times e^{2.12}}{Z_2} = \frac{.3}{Z_2} = .424$

$W_3 = W_4 = \frac{.036 \times e^{2.12}}{Z_2} = \frac{.3}{Z_2} = .424$

$Z_2 = .708$

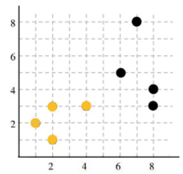
$\varepsilon_3 = \frac{1}{8}[.006 * 1 + .006 * 1] = .0024$

$\alpha_3 = \frac{1}{2}\ln([1 - .0024] / .0024)$

$= 3.01$

3. https://github.com/NootCode/CS4210-Assignment-3/blob/master/Bagging/bagging_random_forest.py

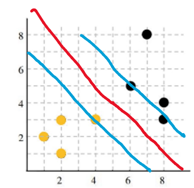
4. [20 points] Say you are given the training dataset shown below. This is a binary classification task in which the instances are described by two integer-valued attributes.



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)



B. No a black circle at (7,5) will not affect the previous decision boundary because it is Not between the support vectors
C. No a yellow circle at (4,2) will not affect the previous decision boundary because it is Not between the support vectors
D. Yes, it will be classified correctly because it is on the correct side of the decision boundary
E. Yes, it will be classified correctly because it is on the correct side of the decision boundary
F. Yes, it will be classified correctly because it is on the correct side of the decision boundary
G. Yes, it will be classified correctly because it is on the correct side of the decision boundary even though it is within the support vector
H. No, it will not be classified correctly because it is on the incorrect side of the DB.
I. No, it will not be classified correctly because it is on the incorrect side of the DB.
J. If C = 1 the penalty will be low so a lot of the sample data will violate the rules of the DB. If C = inf the penalty will be high and will act very similar to a hard margin formulation

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 answers).
b. [4 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. [4 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.

a) The decision boundary lies on the point $x = 0.5$ w/ the support vectors on 0 & 1.

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

$\boxed{b = -1}$

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

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

$\boxed{\text{Margin} = 1}$

$1(w \cdot 1 - 1) = 1$ $\frac{2}{\|w\|}$ Margin $= 1$

$w \cdot 1 = 2$

$\boxed{w = 2}$

c) $-1(w \cdot 0 + b) = 1$ $\boxed{b = -1}$

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

$w \cdot 2 = 2$

$\boxed{w = 1}$

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

Margin $= 2$

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. $\Phi(A)$
b. $\Phi(B)$
c. $\Phi(A)\Phi(B)$
d. $K(A, B)$. Hint: your answers for (c) and (d) should be the same. By using the kernel function, SVM "cheats" and performs significantly fewer calculations. This is known by "kernel trick".

a) $A = (1,2)$

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

b) $B = (2,4)$

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

c) $\phi(A) \cdot \phi(B)$

$= 4 + 64 + 32 + 4 + 16 + 1$

$\boxed{= 121}$

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

$(1,2) \cdot (2,4) = ((2 + 8) + 1)^2$

$= 11^2 = 121$

7. https://github.com/NootCode/CS4210-Assignment-3/blob/master/SVM/svm.py