# CS57800: Statistical Machine Learning

## Homework 4

**Name: Mohammad Haseeb**
**Purdue ID: mhaseeb@purdue.edu**

Due: December 05, 2018 on Wednesday

# 1 Boosting

In this question, $N = 10$.

Initial distribution: $d_n^1 = 1/N = 1/10 = 0.1$

*Note: I calculated the weighted training error for all the hypotheses in each iteration while solving this question on paper. For the sake of saving space and time, I will be showing the working for three hypotheses in each iteration.*

## 1.1 Iteration t = 1

$d_n^1 = 1/N = 1/10 = 0.1$

**Testing different hypotheses**

1. $h$: If $x_1 > 5$ then label +, else label -
   - Indexes of examples labeled incorrectly: 9, 10
   - Error $= 0.1 \times 2 = 0.2$
2. $h$: If $x_2 > 1$ then label -, else label +
   - Indexes of examples labeled incorrectly: 3, 6, 7, 9
   - Error $= 0.1 \times 4 = 0.4$
3. $h$: If $x_2 > 6$ then label +, else label -
   - Indexes of examples labeled incorrectly: 1, 5, 10
   - Error $= 0.1 \times 3 = 0.3$

The hypothesis with the least error is $h$: If $x_1 > 5$ then label +, else label -. Therefore, $h_1 =$ If $x_1 > 5$ then label +, else label -

**Calculate** $\alpha$

$\alpha_1 = \frac{1}{2} ln \frac{1-0.2}{0.2} = 0.69$

**Update weights using** $d_n^{t+1} = d_n^t \cdot exp\{-\alpha_t y_n h_t(x_n)\}/Z_t$

Weights without normalization

| Index | $d_n^2$ |
|---|---|
| 1 | 0.05 |
| 2 | 0.05 |
| 3 | 0.05 |
| 4 | 0.05 |
| 5 | 0.05 |
| 6 | 0.05 |
| 7 | 0.05 |
| 8 | 0.05 |
| 9 | 0.2 |
| 10 | 0.2 |

Table 1: Weights without normalization

$\implies Z_1 = 0.05 * 8 + 0.2 * 2 = 0.8$

Weights with normalization

| Index | $d_n^2$ |
|---|---|
| 1 | 0.0625 |
| 2 | 0.0625 |
| 3 | 0.0625 |
| 4 | 0.0625 |
| 5 | 0.0625 |
| 6 | 0.0625 |
| 7 | 0.0625 |
| 8 | 0.0625 |
| 9 | 0.25 |
| 10 | 0.25 |

Table 2: Weights with normalization

## 1.2 Iteration t = 2

$d_n^2$ are shown in Table 2.

**Testing different hypotheses**

1. $h$: If $x_2 > 9$ then label +, else label -
    - Indexes of examples labeled incorrectly: 1, 3, 5, 6

- Error $= 0.0625 \times 4 = 0.25$

2. $h$: If $x_1 > 6$ then label +, else label -
   - Indexes of examples labeled incorrectly: 9, 10
   - Error $= 0.25 \times 2 = 0.5$
3. $h$: If $x_2 > 4$ then label +, else label -
   - Indexes of examples labeled incorrectly: 1, 4, 5, 10
   - Error $= 0.0625 \times 3 + 0.25 \times 1 = 0.4375$

The hypothesis with the least error is $h$: If $x_2 > 9$ then label +, else label -. Therefore, $h_2 =$ If $x_2 > 9$ then label -, else label +

**Calculate** $\alpha$

$\alpha_2 = \frac{1}{2} ln \frac{1-0.25}{0.25} = 0.55$

**Update weights using** $d_n^{t+1} = d_n^t \cdot exp\{-\alpha_t y_n h_t(x_n)\}/Z_t$

Weights without normalization

| Index | $d_n^3$ |
|---|---|
| 1 | 0.1 |
| 2 | 0.04 |
| 3 | 0.1 |
| 4 | 0.04 |
| 5 | 0.1 |
| 6 | 0.1 |
| 7 | 0.04 |
| 8 | 0.04 |
| 9 | 0.144 |
| 10 | 0.144 |

Table 3: Weights without normalization

$\implies Z_2 = 0.1 \times 4 + 0.04 \times 4 + 0.144 \times 2 = 0.848$

Weights with normalization are shown in Table 4 *(table could not be formatted on this page properly)*.

**Final hypothesis:** $SIGN(0.69(x_1 > 6) + 0.55(x_2 > 9))$

# 2 PAC Learning

## 2.1 Circles

The VC-dimension of circles in the plane $\mathbb{R}^2$ is 3. It is easy to show that any set of points of sizes $K = 1, 2$ can be shattered. For $K = 3$, we can have two cases: (i) where all points are collinear and (ii) when the three points form any other shape. For any combination of labels in both these cases, we can always find a point $p$ and a value for radius $r$ such that we correctly classify the points. As an example, we can have any of the following labels: {- - -, - - +, - + -, - + + , + - - , + - +, + + - , + + + }, then depending on their location on the plane, we can always define a large or small enough circle such that the + points are encompassed by it.

To show that the VC-dimension of circles is not 4, we consider two cases: (i) the points form a triangle such that there are exactly three collinear points , and (ii) one point is inside the convex hull formed by the other three points. For case (i), consider the points (2,2), (3,1), (3,2), (3,3) in the plane $\mathbb{R}^2$. Then, there is no circle of any radius such that these points are labeled correctly. For case (ii), consider the points (2,3), (3,1), (3,2), (4,3) in the plane $\mathbb{R}^2$. Again for this case, there is no circle of any radius such that these points are labeled correctly.

Therefore, the VC-dimension for circles in the plane $\mathbb{R}^2$ is 3.

## 2.2 Triangles

The VC-dimension of triangles in the plane $\mathbb{R}^2$ is 7. It is trivial to show that the sets of points of sizes $K = 1, 2, 3, 4, 5$ can be shattered.

For $K = 6$, consider the points (1,2), (2,1), (2,3), (3,4), (4,1), (5,3). Then, for any labeling of these points we can draw a triangle such that we correctly encompass the positive labeled points. For example, consider the case where we have alternate labeling for the above set of points: +, -, +, -, +, -. Then get three segments of three continuous points such that there is a - label surrounded by two + labels. In such a case, we can dissect each segment near the + labeled points to get a triangle that correctly encompasses the positive points. This can be done for any ordering of the same set of points.

We can also shatter a set of points of size $K = 7$ using a similar argument as above. For example, consider the points (1,2), (2,3), (4,1), (4,4), (5,4), (6,2), (7,3). Again, we get at most three segments of three continuous points such that there is a - label surrounded by two + labels. We can again dissect these segments and get a correct classification of points. This is true for any labeling of the above set of points.

However, it is not possible to shatter a set of points of size $K = 8$. We show this by considering three equivalence classes: (i) collinear points, (ii) one point inside the convex hull formed by the other seven points and (iii), by considering an octagon with alternate labeling of points. For case (i), consider a collinear set of points with the labeling +, -, +, -, +, -, +, -. Then, there is no triangle that can correctly label these points. For case (ii), if the points inside the convex hull is labeled - and all other points are labeled +, then again there is no triangle that can correctly classify this set of points. Finally, for case (iii), consider an octagon consisting of points (1,2), (1,3), (2,1), (2,4), (3,1), (3,4), (4,2), (4,3) labeled alternately as +, -, +, -, +, -, +, -. The only way that this set of points can be shattered is using a quadrilateral.

Therefore, the VC-dimension of triangles is 7.

## 2.3 Trees

(a) If there are $n$ variables then the root can take $n$ different values. Once we have chosen a variable for the root, we are left with $n-1$ variables to choose from that can be the two children of the root (because both the children of the root are required to be the same variable). Finally, because there are two labels (+, -) and four children, there can be a total of $2^4$ possible permutations for the leaves. This gives the following number of syntactically distinct trees:

$$|H_{rd2}| = 2^4 \cdot n \cdot (n-1)$$

(b) Plugging in $|H_{rd2}|$ for $|H|$ into the equation $m > \frac{1}{\epsilon}\{ln(|H|) + ln(\frac{1}{\sigma})\}$, we get the following answer:

$$m > \frac{1}{\epsilon}\{ln(2^4 \cdot n \cdot (n-1)) + ln(\frac{1}{\sigma})\}$$

where $\sigma = 1 - \delta$.

| Index | $d_n^3$ |
|---|---|
| 1 | 0.12 |
| 2 | 0.05 |
| 3 | 0.12 |
| 4 | 0.05 |
| 5 | 0.12 |
| 6 | 0.12 |
| 7 | 0.05 |
| 8 | 0.05 |
| 9 | 0.17 |
| 10 | 0.17 |

Table 4: Weights with normalization