CMSC5724: Exercise List 3

Problem 1. Consider d boolean variables $x_1, x_2, ..., x_d$. An *expression* is a conjunction of d literals where the i-th *literal* ($i \in [1, d]$) is either x_i or its negation $\overline{x_i}$. For example, when $d = 3$, $x_1 \wedge x_2 \wedge \overline{x_3}$ and $\overline{x_1} \wedge x_2 \wedge \overline{x_3}$ are both expressions. Let $\mathcal{H}$ be the set of all expressions. Define the instance space $\mathcal{X} = \{0, 1\}^d$ and label space $\mathcal{Y} = \{-1, 1\}$. Note that each $x \in X$ is a d-dimensional boolean vector $x = (x_1, ..., x_d)$. For each expression (i.e., classifier) $h \in \mathcal{H}$, $h(x)$ equals 1 if evaluating the expression h on x gives 1, or -1 if the evaluation gives 0. Consider $d = 10$. Denote by $\mathcal{D}$ a distribution over $\mathcal{X} \times \mathcal{Y}$. Let S be a training set of objects drawn independently from $\mathcal{D}$. Prove: when $|S| \geq 50000$, with probability at least 0.9, the error of h on $\mathcal{D}$ is higher than the error of h on S by at most 0.01, for every $h \in \mathcal{H}$.

Answer. Let $err_{\mathcal{D}}(h)$ be the error of h on $\mathcal{D}$, and $err_S(h)$ be the error of h on S, for a classifier h in $\mathcal{H}$. By definition, $|\mathcal{H}| = 2^d$. According to the generalization theorem, with probability at least $1 - \delta$, we have

$$err_{\mathcal{D}}(h) \quad \leq \quad err_S(h) + \sqrt{\frac{\ln(1/\delta) + \ln|\mathcal{H}|}{2|S|}}$$

for every $h \in H$. By setting $\delta = 0.1$, we know with probability at least 0.9,

$$err_{\mathcal{D}}(h) \quad \leq \quad err_S(h) + \sqrt{\frac{\ln(10 \cdot 2^{10})}{2|S|}}.$$

When $|S| \geq 50000$, we have $\sqrt{\frac{\ln(10 \cdot 2^{10})}{2|S|}} \leq 0.01$; hence, with probability at least 0.9, $err_{\mathcal{D}}(h)$ is higher than $err_S(h)$ by at most 0.01, for every $h \in \mathcal{H}$.

Problem 2. Let P be a set of 4 points: $A = (1, 2, 1)$, $B = (2, 1, 1)$, $C = (0, 1, 1)$ and $D = (1, 0, 1)$. A and B have label 1, while C and D have label -1. Execute Perceptron on P. Give the weight vector $\boldsymbol{w}$ maintained by the algorithm after each iteration.

Answer. At the beginning, $\boldsymbol{w} = (0, 0, 0)$. We use $\boldsymbol{A}$ to denote the vector form of A. Define $\boldsymbol{B}, \boldsymbol{C}$, and $\boldsymbol{D}$ similarly.

Iteration 1. Since $\boldsymbol{A}$ does not satisfy $\boldsymbol{A} \cdot \boldsymbol{w} > 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} + \boldsymbol{A} = (0, 0, 0) + (1, 2, 1) = (1, 2, 1)$.

Iteration 2. Since $\boldsymbol{C}$ does not satisfy $\boldsymbol{C} \cdot \boldsymbol{w} < 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} - \boldsymbol{C} = (1, 2, 1) - (0, 1, 1) = (1, 1, 0)$.

Iteration 3. Since $\boldsymbol{C}$ does not satisfy $\boldsymbol{C} \cdot \boldsymbol{w} < 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} - \boldsymbol{C} = (1, 1, 0) - (0, 1, 1) = (1, 0, -1)$.

Iteration 4. Since $\boldsymbol{A}$ does not satisfy $\boldsymbol{A} \cdot \boldsymbol{w} > 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} + \boldsymbol{A} = (1, 0, -1) + (1, 2, 1) = (2, 2, 0)$.

Iteration 5. Since $\boldsymbol{C}$ does not satisfy $\boldsymbol{C} \cdot \boldsymbol{w} < 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} - \boldsymbol{C} = (2, 2, 0) - (0, 1, 1) = (2, 1, -1)$.

Iteration 6. Since $\boldsymbol{C}$ does not satisfy $\boldsymbol{C} \cdot \boldsymbol{w} < 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} - \boldsymbol{C} = (2, 1, -1) - (0, 1, 1) = (2, 0, -2)$.

Iteration 7. Since $\boldsymbol{A}$ does not satisfy $\boldsymbol{A} \cdot \boldsymbol{w} > 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} + \boldsymbol{A} = (2, 0, -2) + (1, 2, 1) = (3, 2, -1)$.

Iteration 8. Since $\boldsymbol{C}$ does not satisfy $\boldsymbol{C} \cdot \boldsymbol{w} < 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} - \boldsymbol{C} = (3, 2, -1) - (0, 1, 1) = (3, 1, -2)$.

Iteration 9. Since $\boldsymbol{D}$ does not satisfy $\boldsymbol{D} \cdot \boldsymbol{w} < 0$, update $\boldsymbol{w}$ to $\boldsymbol{w} - \boldsymbol{D} = (3, 1, -2) - (1, 0, 1) = (2, 1, -3)$.

Iteration 10. No more violation points.

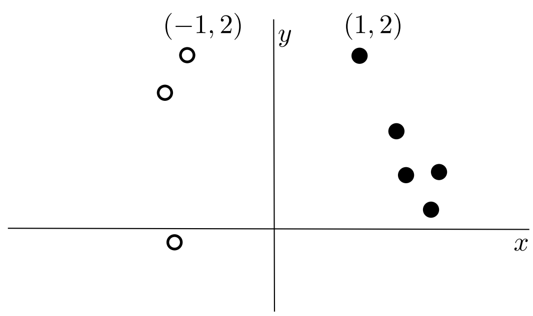
Problem 3. Let P be a set of multidimensional points where each point has a label equal to 1 or -1. We want to design an algorithm to achieve the following purpose:

- Either return a separation plane (see the lecture notes for the definition of separation plane);
- Or declare that P has no separation planes with a margin at least γ.

Your algorithm must still work even if no separation planes exist.

Answer. Run Perceptron and return whatever plane found by the algorithm. If the algorithm still has not finished after R^2/γ^2 corrections, force it to stop and declare that no separation plane has a margin at least γ.

Problem 4. Consider the set of points below where points of different colors carry different labels. Only two points have their coordinates shown. Apply Perceptron to find a separation plane on the set. Prove: Perceptron finishes after at most 5 iterations.



Answer. The y-axis is a separation plane with margin $\gamma = 1$. Clearly, the largest distance from a point to the origin is $R = \sqrt{5}$. Hence, Perceptron performs at most $R^2/\gamma^2 = 5$ iterations.

Problem 5. Some people prefer the following variant of the Perceptron algorithm:

```
w = 0
while there is a violating point p
    if p has label 1
        w = w + λ · p
    else
        w = w − λ · p
```

where λ is a positive real-value constant. In the version we discussed in the lecture, $\lambda = 1$. Prove: regardless of λ, Perceptron always terminates in R^2/γ^2 iterations, where R is the maximum distance of the points to the origin and γ the largest margin of all separation planes.

Answer. Let $\boldsymbol{w}_k$ denote the vector $\boldsymbol{w}$ after the k-th iteration. Following the analysis we discussed in the lecture, we can prove the two inequalities below:

$$\begin{aligned} |\boldsymbol{w}_k| &\geq \lambda \cdot k \cdot \gamma \\ |\boldsymbol{w}_k|^2 &\leq k \cdot \lambda^2 \cdot R^2 \end{aligned}$$

The two inequalities give $\lambda^2 \cdot k^2 \cdot \gamma^2 \leq k \cdot \lambda^2 \cdot R^2$, which indicates $k \leq R^2/\gamma^2$.

Problem 6. Let P be a set of points in $\mathbb{R}^d$, where each point is labeled 1 or -1. A d-dimensional plane π is a *separation plane* of P if

- π does not pass any point in P;
- the points of the two labels in P fall on different sides of π.

Note that we do not require π to pass the origin.

Construct a $(d+1)$-dimensional point set P' as follows: given each $p \in P$, add to P' the point $(p[1], p[2], ..., p[d], 1)$ (i.e., adding a new coordinate 1), carrying the same label as p. Prove: P has a separation plane if and only if P' has a separation plane passing the origin of $\mathbb{R}^{d+1}$.

Answer: Given a point $p \in P$, we use $\boldsymbol{p}$ to denote its vector form. Similarly, we use $\boldsymbol{p'}$ to denote the vector form of a point $p' \in P'$.

Only-if direction. Suppose that P has a separation plane. Then, there must be a d-dimensional plane $\boldsymbol{w} \cdot \boldsymbol{x} + w_{d+1} = 0$ such that for every $p \in P$:

$$\begin{cases} \boldsymbol{w} \cdot \boldsymbol{p} + w_{d+1} > 0 & \text{if } p \text{ has label } 1 \\ \boldsymbol{w} \cdot \boldsymbol{p} + w_{d+1} < 0 & \text{if } p \text{ has label } -1. \end{cases} \tag{1}$$

Let $\boldsymbol{w'} = (\boldsymbol{w}[1], \boldsymbol{w}[2], ..., \boldsymbol{w}[d], w_{d+1})$. Every $p' \in P'$ has the same label as $(p'[1], p'[2], ..., p'[d])$ in P and $p'[d+1] = 1$. From (1), we have

- $\boldsymbol{w'} \cdot \boldsymbol{p'} = \sum_{i=1}^{d} \boldsymbol{w}[i] p'[i] + w_{d+1} > 0$ if p' has label 1;
- $\boldsymbol{w'} \cdot \boldsymbol{p'} = \sum_{i=1}^{d} \boldsymbol{w}[i] p'[i] + w_{d+1} < 0$ if p' has label -1.

Hence, P' has a separation plane (i.e., $\boldsymbol{w'} \cdot \boldsymbol{x} = 0$) passing the origin of $\mathbb{R}^{d+1}$.

If-direction. Suppose that P' has a separation plane passing the origin of $\mathbb{R}^{d+1}$. Then, there must be a $(d+1)$-dimensional vector $\boldsymbol{w'}$ such that for every $p' \in P'$:

$$\begin{cases} \boldsymbol{w'} \cdot \boldsymbol{p'} > 0 & \text{if } p' \text{ has label } 1 \\ \boldsymbol{w'} \cdot \boldsymbol{p'} < 0 & \text{if } p' \text{ has label } -1. \end{cases} \tag{2}$$

Let $\boldsymbol{w} = (\boldsymbol{w'}[1], \boldsymbol{w'}[2], ..., \boldsymbol{w'}[d])$. Every $p \in P$ has the same label as $p' = (p[1], p[2], ..., p[d], 1)$ in P'. From (2), we know

- $\boldsymbol{w} \cdot \boldsymbol{p} + \boldsymbol{w'}[d+1] = \boldsymbol{w'} \cdot \boldsymbol{p'} > 0$ if p has label 1;
- $\boldsymbol{w} \cdot \boldsymbol{p} + \boldsymbol{w'}[d+1] = \boldsymbol{w'} \cdot \boldsymbol{p'} < 0$ if p has label -1.

It thus follows that P has a separation plane $\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{w'}[d+1] = 0$.