CS6140: Machine Learning – Homework 5
Name: Vishal Rajpal

Problem 1: Classification via Linear Programming

a) Since, we already did it in class, I have understood most of it and I did it in a similar way.

Hard

$z =$

s.t.

$$S = \{(\bar{x}_1, y_1), \ldots, (x_m, y_m)\}$$

$$x_i \in R^d \; ; \; y \in \{-1, 1\}$$

$$h(\bar{x}_1) = 1 \quad \text{if} \quad \bar{w} \cdot \bar{x}_1 + \theta \geq 0$$

else -1

$$\min \quad z = \xi$$

$$\text{s.t.} \quad y_1(\bar{w}\,\bar{x} + \theta) \geq 1 - \xi \qquad \forall (x_1, y_1) \in S$$

$$\xi \geq 0$$

a) show that linearly separable implies $\xi = 0$

$$\max_{\substack{(\bar{x}, \bar{y}) \in S \\ y = -1}} \bar{v} \cdot \bar{x} + \delta < 0 \leq \min_{\substack{(x, y) \in S \\ y = 1}} \bar{v} \cdot \bar{x} + \delta$$

$\bar{x}^+$ position instance to $\bar{v} \cdot \bar{x} + \delta$

$$\text{s.t.} \quad p^+ = \bar{v} \cdot \bar{x}^+ + \delta$$

$\bar{x}^-$ - negative instance.

$$p^- = \bar{v} \cdot \bar{x}^- + \delta$$

$$p^+ > p^-$$

Specifically consider hyperplane defined by y

$$\bar{v} \cdot \bar{x}^+ + \delta - \gamma$$

noting that

$$\bar{v} \cdot \bar{x}^+ + \delta - \gamma = -(\bar{v} \cdot \bar{x}^- + \delta - \gamma)$$

$$\therefore \gamma = \frac{p^+ + p^-}{2}$$

$$p^- - \gamma < 0 \leq p^+ - \gamma$$

Negative hyperplane

Slack var
allows t
cases.

$$\bar{v} \cdot \bar{x}^{\circ} + \delta - \gamma = \frac{p^+ - p^-}{2}$$

and

$$\bar{v} \cdot \bar{x}^- + \delta - \gamma = \frac{p^+ - p^-}{2}$$

$$y(\bar{v} \cdot \bar{x}^{\circ} + \delta - \gamma) \geq \frac{p^+ - p^-}{2} \quad \forall (x, y) \in S$$

Given that $p^+ > p^-$ and $\gamma = \frac{p^+ + p^-}{2}$

$$\bar{\omega} = \frac{\bar{v}}{\gamma}, \quad \theta = \frac{\delta - \gamma}{\gamma} \quad \text{and} \quad \xi > 0$$

If exist $\bar{\omega} \cdot \bar{x} + \theta$ st.

$$y(\bar{\omega} \cdot \bar{x} + \theta) \geq 1 \quad \forall (x, y) \in S$$

then

$$\bar{\omega} \cdot \bar{x} + \theta \geq 1 \; \forall (x, y) \in S \leftarrow y = 1$$

and

$$\bar{\omega} \cdot \bar{x} + \theta \leq -1$$

$$\forall (x, y) \in S \leftarrow y = -1$$

b) Linear Program

$$c = [0^{d+1} 1]^T$$
$$x = [w\ \theta\ \xi]^T$$

$$A = \begin{matrix} y_1 x_1^T & y_1 & 1 \\ \vdots & \vdots & \vdots \\ y_m x_m^T & y_m & 1 \\ 0^d & 0 & 1 \end{matrix}$$

$$b = \left[1^{[d+1]}\ 0\right]^T$$

Problem 2: VC Dimensions

a) Considering the case of a single point. Based on our hypotheses $||x|| < a$, we can determine that a single point can be easily shattered such that if the point is positive then $||x|| < a$ and if it is negative $||x|| \geq a$.

Considering the case of two points. Consider two points, $||x_1|| \leq ||x_2||$ if $x_1$is positive and x_2 is negative, then these points cannot be shattered. Therefore, $VC = 1$

b) Considering the case of a single point. Based on our hypotheses $a < \ ||x|| < b$, a single point can easily be shattered such if the point is positive then a and b can be chosen based on the hypotheses and even if the point is negative, $a\ and\ b$ can be chosen such that the point lies outside the bagel.

Considering the case of two points. If we have two points such that $||x_1|| < ||x_2||$
- If both the points are positive, then we can have $a < ||x_1||$ and $b > ||x_2||$.
- If both points are negative, then we can have $a < b < ||x_1||$ or $b > a > ||x_2||$.
- If point closer to origin i.e. $||x_1||$ is positive and $||x_2||$ is negative, we can have $a < ||x_1|| < b < ||x_2||$.
- If point closer to origin i.e. $||x_1||$ is negative and $||x_2||$ is positive, we can have $||x_1|| < a < ||x_2|| < b$.

Considering the case of three points, there are two cases where we cannot shatter them, Consider three points $||x_1|| < \ ||x_2|| < \ ||x_3||$
- If $||x_1||$ and $||x_3||$ are positive and $||x_2||$ is negative, then we cannot shatter these points.
- If any of the points are equidistant from origin and if we label them differently, then cannot be shattered.

Therefore, VC $= 2$

c) Considering the case of a single point, we can easily shatter it based on having a point inside the circle if it is positive and outside, if it is negative.
Considering the case of two points, again we can have a positive point inside the circle and negative points outside the circle. Two points are can be easily shattered.

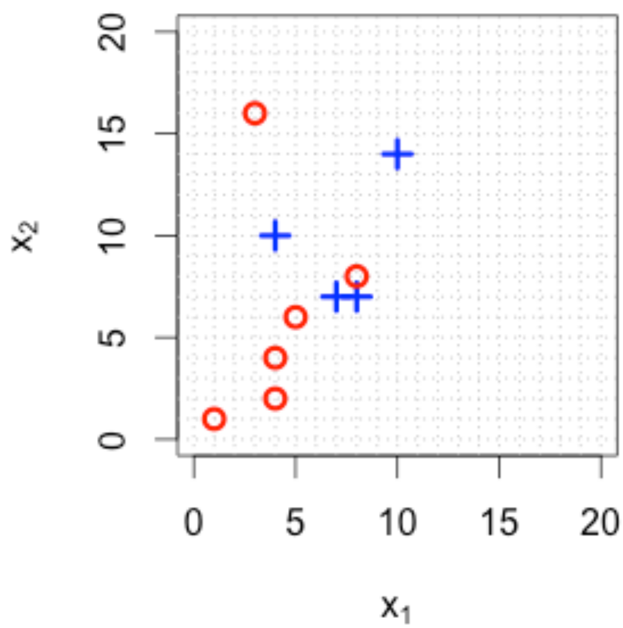
Considering the case of three points
- If no points are positive, the circle is empty.
- If all points are positive, the circle contains all the points.
- If one point is positive and two points are negative, we can have a circle surrounding that positive point.
- If one point is negative, and two are positive we can still have a circle surrounding just the positive points can be shattered except a case when the points are collinear and the outer middle point is negative and the other points are labeled positive.

Considering the case of 4 points

- Consider a use case similar to XOR where these points are labeled same cross the lines and are equidistant from each other, they cannot be shattered.

Therefore, $VC = 3$

Problem 3: Adaboost



		Hypothesis 1			Hypothesis 2		
id	y	D_0	$h_{x1} = [x_1 > 6]$	$h_{x2} = [x_2 > 6]$	D_1	$h_{x1} = [x_1 > 8]$	$h_{x2} = [x_2 > 8]$
1	-	$\frac{1}{10}$	-	-	$\frac{1}{16}$	-	-
2	-	$\frac{1}{10}$	-	+	$\frac{1}{4}$	-	+
3	-	$\frac{1}{10}$	-	-	$\frac{1}{16}$	-	-
4	-	$\frac{1}{10}$	+	+	$\frac{1}{4}$	-	-
5	-	$\frac{1}{10}$	-	-	$\frac{1}{16}$	-	-
6	-	$\frac{1}{10}$	-	-	$\frac{1}{16}$	-	-
7	+	$\frac{1}{10}$	+	+	$\frac{1}{16}$	-	-
8	+	$\frac{1}{10}$	-	+	$\frac{1}{16}$	-	+
9	+	$\frac{1}{10}$	+	+	$\frac{1}{16}$	+	+
10	+	$\frac{1}{10}$	+	+	$\frac{1}{16}$	-	-
$error = \varepsilon$			0.2	**0.2**		0.3	0.3

a) See above
b) See above table
c) See above table. The error in both the hypotheses is 0.2 and we choose $h_{x2} = [x_2 > 6]$. Therefore $\alpha_0 = \frac{1}{2}\log\left[\frac{[1-0.2]}{0.2}\right] = \frac{1}{2}\log\left[\frac{0.8}{0.2}\right] = \frac{1}{2}\log(4) = 1$
d) $D_1 = \frac{1}{2}\left[\frac{D_0}{Z_0}\right]\ if\ \ h_{1x2} = y_i\ , 2\left[\frac{D_0}{Z_0}\right]\ if\ \ h_{1x2} \neq y_i$

$$D_1 = \frac{1}{20Z_0}\ if\ \ h_{1x2} = y_i\ , \frac{1}{5Z_0}\ \ if\ \ h_{1x2} \neq y_i$$

$Based\ on\ the\ number\ of\ correct\ and\ incorrect\ examples$

$$\frac{8}{20Z_0} + \frac{2}{5Z_0} = 1$$

$$\frac{80}{100Z_0} = 1$$

$$Z_0 = \frac{4}{5}$$

Therefore,

$$D_1 = \frac{1}{20 * \left(\frac{4}{5}\right)}\ if\ \ h_{1x2} = y_i\ , \frac{1}{5\left(\frac{4}{5}\right)}\ \ if\ \ h_{1x2} \neq y_i$$

$$D_1 = \frac{1}{16}\ if\ \ h_{1x2} = y_i\ , \frac{1}{4}\ \ if\ \ h_{1x2} \neq y_i$$

e) Final Hypothesis
After second round

$$\varepsilon_{x1} = \frac{3}{16}\ , \varepsilon_{x2} = \frac{1}{4} + \frac{2}{16} = \frac{6}{16} = \frac{3}{8}$$

Therefore,

$$\alpha_1 = \frac{1}{2}\log\left[\frac{\left[1 - \frac{3}{16}\right]}{\frac{3}{16}}\right] = \frac{1}{2}\log\left[\frac{\frac{13}{16}}{\frac{3}{16}}\right] = \frac{1}{2}\log(4.33) = 1.06$$

$$\boldsymbol{H(x) = [x_2 > 6] + 1.06[x_2 > 8]}$$