ASSIGNMENT 3, Solutions

1 (200pts) Coin Model / Bin Model

(a) We want

$$\begin{aligned} P[\nu = 0 \text{ for at least 1 coin}] &= 1 - P[\nu \neq 0 \text{ for all bins}] \\ &= 1 - P[\nu_1 \neq 0]^M \\ &= 1 - (1 - (1-\pi)^N)^M \end{aligned}$$

where N is number of tosses and M the number of coins. We have $N = 10$, thus plugging in $M = \{1, 1000, 10^6\}$ and $\pi = \{0.05, 0.8\}$ we get

	M		
pi	1	1000	10^6
0.05	0.599	$1 - 2.7 \times 10^{-397} \approx 1$	$\geq 1 - 10^{-50000} \approx 1$
0.8	10^{-7}	10^{-4}	9.7×10^{-2}

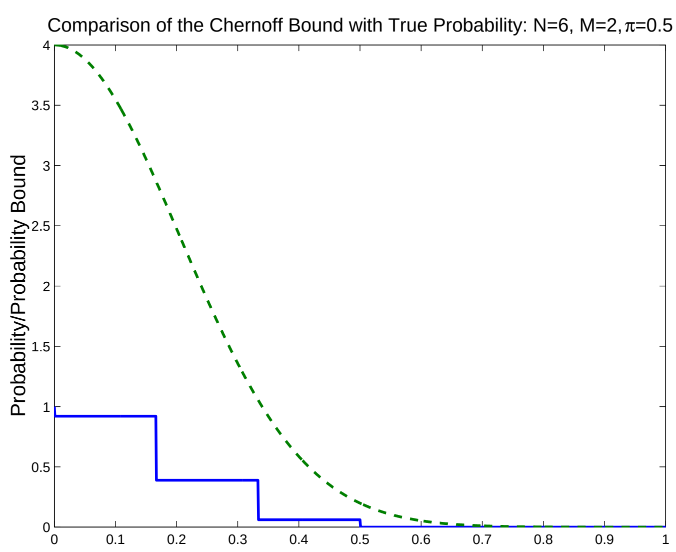
(b) ν can take on one of the seven values $\{\frac{0}{6}, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, \frac{6}{6}\}$ with $P[\nu = \frac{i}{6}] = \binom{6}{i}(\frac{1}{2})^6$. Thus

ϵ	0	$0 < \epsilon \leq \frac{1}{6}$	$\frac{1}{6} < \epsilon \leq \frac{2}{6}$	$\frac{2}{6} < \epsilon \leq \frac{3}{6}$	$\frac{3}{6} < \epsilon$
$P[\lvert\nu - \pi\rvert \geq \epsilon]$	1	$\frac{22}{32}$	$\frac{7}{32}$	$\frac{1}{32}$	0

$P[\max_i \lvert\nu_i - \pi_i\rvert \geq \epsilon] = P[\lvert\nu_1 - \pi_1\rvert \geq \epsilon \text{ or } \lvert\nu_1 - \pi_1\rvert \geq \epsilon] = 2P[\lvert\nu - \pi\rvert \geq \epsilon] - P[\lvert\nu - \pi\rvert \geq \epsilon]^2$, so we get

ϵ	0	$0 < \epsilon \leq \frac{1}{6}$	$\frac{1}{6} < \epsilon \leq \frac{2}{6}$	$\frac{2}{6} < \epsilon \leq \frac{3}{6}$	$\frac{3}{6} < \epsilon$
$P[\max_i \lvert\nu_i - \pi_i\rvert \geq \epsilon]$	1	$\frac{231}{256}$	$\frac{399}{1024}$	$\frac{63}{1024}$	0

The Chernoff bound for 2 coins gives $P[\max_i \lvert\nu_i - \pi_i\rvert \geq \epsilon] \leq 4exp(-2N\epsilon^2)$. The true probability and the Chernoff bound are plotted below.



2 (200pts) Computation of $M(N)$

(a) Let the left end point of the interval be l and the right be r and suppose that the points are ordered, $x_1 < x_2 < \ldots < x_N$ with $N \geq 3$. There are $\binom{N-1}{2}$ ways of placing both end points in different intervals chosen from the $N-1$ intervals between x_1 and x_N. Each of these choices yields 2 dichotomies (positive or negative interval). If $l < x_1$ then there are N ways of placing r in a different interval, once again each of these N ways yields 2 dichotomies. There are no other dichotomies therefore

$$m(N) = 2\binom{N-1}{2} + 2N = N^2 - N + 2$$

This formula also holds for $N = 1$ and $N = 2$. $m(3) = 8$ and $m(4) = 14$ so $d_{VC} = 3$.

(b) Let the N points be $x_1, \ldots, x_N$, with $\| x_1 \| < \| x_2 \| < \ldots < \| x_N \|$. Putting these on a line, we see that this learning model is equivalent to the positive interval model considered in class, so

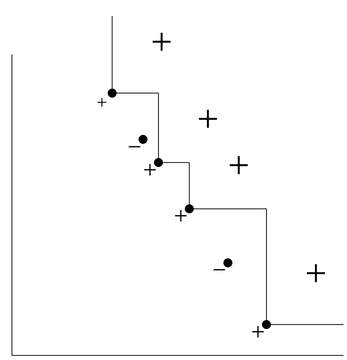
$$m(N) = 1 + \binom{N+1}{2} = 1 + \frac{N(N+1)}{2}$$

$m(2) = 4$ and $m(3) = 7$ so $d_{VC} = 2$.

3 (400pts) VC Dimension

(a) Choose N different points in $\mathbf{R}^2$ to be $(i, 0)$. Given a dichotomy, there is a polynomial of order N in x that takes on the value of this dichotomy. Thus taking the sign will not affect the value, hence we can implement any dichotomy (as the polynomials are continuous). Thus $d_{VC} = \infty$.

(b) Consider the points arranged along the $y = 1 - x$ line as shown in the figure



Given any dichotomy, for example the one shown in the figure, let $\mathbf{x}_i^+$ be the points with $f(x) = +1$. The function

$$g(\mathbf{x}) = \begin{cases} 1 \text{ if } \mathbf{x} \geq \mathbf{x}_i^+ \text{ for some } \mathbf{x}_i^+ \text{ in the dichotomy} \\ -1 \text{ otherwise} \end{cases}$$

The decision boundary is shown in the figure. $g(\mathbf{x})$ is monotonic, so we have shown how to implement any dichotomy. Thus, $d_{VC} = \infty$.

(c) (i) Choose the $d+1$ points and define the $(d+1) \times (d+1)$ matrix $\mathbf{Y}$ as follows.

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{x}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \ \mathbf{x}_{d+1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_{d+1} \end{bmatrix}$$

By construction, $Det(\mathbf{Y}) \neq 0$ so $\mathbf{Y}$ is invertible. Given any dichotomy δ, we want to find a $\tilde{\mathbf{w}}$ such that $\delta_i = sign(\tilde{\mathbf{w}}^T \tilde{x}_i)$, where $\tilde{\mathbf{x}}_i = [1 \ \mathbf{x}_i^T]^T$. It suffices to find a $\tilde{\mathbf{w}}$ such that $\delta_i = \tilde{\mathbf{w}}^T \tilde{x}_i$, or in matrix form

$$\delta = \mathbf{Y}^T \tilde{\mathbf{w}} \Rightarrow \tilde{\mathbf{w}} = (\mathbf{Y}^T)^{-1} \delta$$

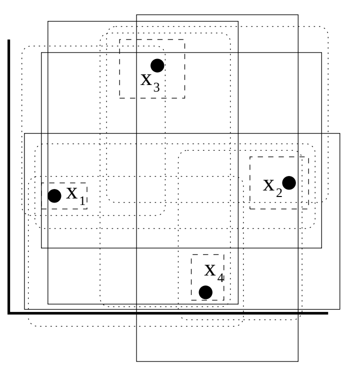
Since $\mathbf{Y}$ is invertible, then so is $\mathbf{Y}^T$, hence, such a $\tilde{\mathbf{w}}$ exists, and, since δ was arbitrary, we have shown that for any dichotomy on these $d+1$ points, a perceptron can implement it. Thus $d_{VC} \geq d+1$.

(ii) Let $\mathbf{x}_1, \ldots, \mathbf{x}_{d+2}$ be any set of $d+2$ vectors. Then $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_{d+2}$ are $d+2$ vectors in $d+1$ dimensions, hence they must be linearly dependent. Thus, for some β, $\tilde{\mathbf{x}}_\beta = \sum_{i \neq \beta} a_i \tilde{\mathbf{x}}_i$. Consider the following dichotomy: $\delta_i = sign(a_i)$ for $i \neq \beta$ and $\delta_\beta = -1$. We prove that this dichotomy cannot be implemented as follows. Suppose that the weight vector $\mathbf{w}$ implements the dichotomy for $i \neq \beta$. Then, for $i \neq \beta$, $\delta_i = sign(a_i) = sign(\mathbf{w}^T \tilde{\mathbf{x}}_i)$ thus a_i and $\mathbf{w}^T \tilde{\mathbf{x}}_i$ have the same sign so $a_i \mathbf{w}^T \tilde{\mathbf{x}}_i \geq 0$. Now consider $\tilde{\mathbf{x}}_\beta$.

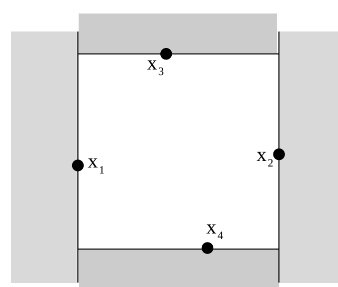
$$\mathbf{w}^T \tilde{\mathbf{x}}_\beta = \sum_{i \neq \beta} a_i \mathbf{w}^T \tilde{\mathbf{x}}_i \geq 0$$

Thus, this weight vector cannot implement $\delta_\beta = -1$, so no vector can simultaneously implement the dichotomy $\delta_i = sign(a_i)$ for $i \neq \beta$ and $\delta_\beta = -1$. Thus on any set of $d+2$ points, there is at least one dichotomy that cannot be implemented. Thus $d_{VC} < d+2$.

(d) $d_{VC} = 4$ (see below) hence $m(N) \leq N^4 + 1$.



(a)



(b)

(i) Consider the 4 points as arranged in figure (a). It is easy to see that any of the 2^4 dichotomies can be implemented with the positive rectangle learning model. Thus, $d_{VC} \geq 4$.

(ii) Of the 5 points, let $\mathbf{x}_1$ be the leftmost and $\mathbf{x}_2$ the rightmost. Thus the remaining points are in the strip between $\mathbf{x}_1$ and $\mathbf{x}_2$. Let $\mathbf{x}_3$ be the topmost of these and $\mathbf{x}_4$ be the bottommost. $\mathbf{x}_3$ must be above both $\mathbf{x}_1$ and $\mathbf{x}_2$ and $\mathbf{x}_4$ must be below both $\mathbf{x}_1$ and x_2. Any other arrangement will leave one point in the smallest rectangle containing two other points and thus there will be a dichotomy that cannot be implemented. The situation is shown in figure (b), the shading indicating no points allowed. The 5^{th} point can now only be placed in the middle and now the dichotomy with $\delta_5 = -1$ and $\delta_1 = \delta_2 = \delta_3 = \delta_4 = +1$ cannot be implemented. Thus we have shown that on any set of 5 points, there is a dichotomy that cannot be implemented. Thus $d_{VC} < 5$.

4 (200pts) Bounds on the Risk

(a) We need to find N such that $6m(2N)e^{-\epsilon^2 N} \leq \eta$ where $\epsilon = 0.05$ and $\eta = 0.05$. One could use $m(N) \leq N^{d_{VC}} + 1$ or $m(N) \leq \left(\frac{eN}{d_{VC}}\right)^{d_{VC}}$ to get a lower bound for N so that this inequality holds.

d_{VC}	5	6
$m(N) \leq N^{d_{VC}} + 1$	$N \geq 23,425$	$N \geq 28,169$
$m(N) \leq \left(\frac{eN}{d_{VC}}\right)^{d_{VC}}$	$N \geq 23,425$	$N \geq 28,169$

(b) $\frac{\pi-\nu}{\sqrt{\pi}} \geq \epsilon$ with probability$\leq \eta = Cm(2N)e^{-\epsilon^2 N/4}$. Solving for ϵ, we find that $\epsilon^2 = \xi$. $\frac{\pi-\nu}{\sqrt{\pi}} < \epsilon$ with probability$\geq 1-\eta$ so,

$$\begin{aligned} \pi &< \nu + \epsilon\sqrt{\pi} < \nu + \epsilon\sqrt{\nu + \epsilon\sqrt{\pi}} \\ &< \nu + \epsilon\sqrt{\nu + \epsilon\sqrt{\nu + \epsilon\sqrt{\ldots}}} \end{aligned}$$

thus $\pi < \nu + \epsilon S$ where $S \geq 0$ satisfies $S^2 = \nu + \epsilon S$. Solving for S and choosing the positive root we see that $S = \frac{\epsilon}{2}\left(1 + \sqrt{1 + 4\nu/\epsilon^2}\right)$.Thus

$$\pi < \nu + \frac{\epsilon^2}{2}\left(1 + \sqrt{1 + \frac{4\nu}{\epsilon^2}}\right) = \nu + \frac{\xi}{2}\left(1 + \sqrt{1 + \frac{4\nu}{\xi}}\right)$$

Alternate Approach:

$$\pi - \epsilon\sqrt{\pi} - \nu < 0 \Rightarrow (\sqrt{\pi} - a_+)(\sqrt{\pi} - a_-) < 0$$

where $a_\pm = \frac{\epsilon}{2}\left(1 \pm \sqrt{1 + 4\nu/\epsilon^2}\right)$. Thus $a_- < \sqrt{\pi} < a_+$ but since $0 \leq a_-$, we get that $0 < \sqrt{\pi} < a_+$ and thus that $\pi < a_+^2$ which then yields the desired result since

$$a_+^2 = \frac{\epsilon^2}{4}\left(1 + 2\sqrt{1 + \frac{4\nu}{\epsilon^2}} + 1 + \frac{4\nu}{\epsilon^2}\right) = \nu + \frac{\epsilon^2}{2}\left(1 + \sqrt{1 + \frac{4\nu}{\epsilon^2}}\right)$$