CSCI 6100 Machine Learning From Data
Fall 2018

HOMEWORK 4
Daniel Southwick
661542908
southd@rpi.edu

Exercise 2.4

(a) Let

$$\mathbb{X} = \begin{bmatrix} x_{10} & x_{11} & \dots & x_{1d} \\ x_{20} & x_{21} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d+10} & x_{d+11} & \dots & x_{d+1d} \end{bmatrix} \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{d+1} \end{bmatrix}$$

Where $\mathbb{X}$ is a nonsingular matrix with dimension $(d+1) \times (d+1)$ whose rows represents the $d+1$ points and columns represents the individual elements of a x. Note that $y = [\pm 1, \pm 1, ..., \pm 1]^T$

If the system $sign(\mathbb{X}w) = y$ has a solution w, then the perceptron can shatter $(d+1)$ points. And $\mathbb{X}$ is a non singular matrix so $\mathbb{X}^{-1}$ can be computed thus $w = \mathbb{X}^{-1}y$. A solution of w is found, so $d_{vc} \geq d+1$.

(b) Same as part (a), we consider each points of $\mathbb{X}$ as $[x_0, x_1, ..., x_d]$. Then any new vector should be the linear combination of the $d+1$ vector. So any $d+2$ vectors of length $d+1$ have to be linearly dependent.

Problem 2.3

(a)Positive or negative ray: For N data points, the line is split by a single point into $N+1$ regions. So positive rays $m_H(N)$ is $N+1$. And same with the negative, thus the maximal number of dichotomies is $2(N+1)-2 = 2N$ (Subtracting two cases where all points are +1 or -1). So $m_H(N) = 2N$. $m_{\mathcal{H}}(3) = 6 < 2^3,\ m_H(2) = 4 = 2^2$ Thus $d_{vc} = 2$

(b)Positive or negative intervals: For $N+1$ regions, If the interval covers the part of two end regions, then the number of dichotomies is $2N$ like part (a), if the intervals lies within the middle $N-1$ regions, then the number of dichotomies became $2\binom{N+1}{2} = (N-1)(N-2)$. Thus the maximum dichotomies: $m_H(N) = (N-1)(N-2) + 2N = N^2 - N + 2$. $m_H(4) = 14 < 2^4,\ m_{\mathcal{H}}(3) = 8 = 2^3$ Thus $d_{vc} = 3$

(c)Two concentric spheres in $\mathbb{R}^d$: The problem is essentially the same as finding the interval. We can map a point in $\mathbb{R}^d$ into a point $y \in \mathbb{R}$ by $y = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$. So for any N points in $\mathbb{R}^d$, we can find the corresponding $y \in \mathbb{R}$ and then it's splits to at most $N+1$

regions, so $m_{\mathcal{H}}(N) = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$. $m_H(3) = 7 < 2^3$, $m_H(2) = 4 = 2^2$. Thus $d_{vc} = 2$

Exercise 2.8

For any growth function that has a break point, it then can be bounded by a polynomial function of N. If not, then $m_H(N) = 2^N$. So $1+N$, $1+N+\frac{N(N-1)}{2}$, 2^N are possible growth functions.
For $m_H(N) = 1+N$: $m_H(1) = 2 = 2^1$ and $m_H(2) = 3 < 2^2 d_{vc} = 1 m_H(N) \leq \sum_{i=0}^{1}\binom{N}{i} = N+1$.
For $m_H(N) = 1+N+\frac{N(N-1)}{2}$: $m_H(2) = 4 = 2^2$ and $m_H(3) = 7 < 2^3 d_{vc} = 2 m_H(N) \leq \sum_{i=0}^{2}\binom{N}{i} = \frac{N(N-1)}{2} + N + 1$.
For $m_H(N) = 2^N$: $m_H(N) = 2^N d_{vc} = \infty m_H(N) \leq \sum_{i=0}^{N}\binom{N}{i} = 2^N$

And $2^{\lfloor\sqrt{N}\rfloor}$, $m_H(N) = 2^{\lfloor N/2\rfloor}$ and $m_H(N) = 1+N+\frac{N(N-1)(N-2)}{6}$ are not possible growth functions
For $m_H(N) = 2^{\lfloor\sqrt{N}\rfloor}$: $m_H(1) = 2 = 2^1$ and $m_H(2) = 2 < 2^2 d_{vc} = 1 m_H(N) \nleq \sum_{i=0}^{1}\binom{N}{i} = N+1$
For $m_H(N) = 2^{\lfloor N/2\rfloor}$: $m_H(0) = 1 = 2^0$ and $m_H(1) = 1 < 2^1 d_{vc} = 0 m_H(N) \nleq \sum_{i=0}^{0}\binom{N}{i} = 1$
For $m_H(N) = 1+N+\frac{N(N-1)(N-2)}{6}$: $m_H(1) = 2 = 2^1$ and $m_H(2) = 3 < 2^2 d_{vc} = 1 m_H(N) \nleq \sum_{i=0}^{1}\binom{N}{i} = N+1$

Problem 2.10

We first separate the group of $2N$ points into two groups of N points, we know that for both groups, their maximal number of dichotomies is $m_H(N)$. So for the whole $2N$ points, the maximal number of dichotomies $m_H(2N)$ are at most $m_H(N) \times m_H(N) = m_H(N)^2$, thus $m_H(2N) \leq m_H(N)^2$

Problem 2.12

We know that the sampling bound $= \sqrt{\frac{8}{N}\ln\frac{4m_{\mathcal{H}}(2N)}{\delta}} = \sqrt{\frac{8}{N}\ln\frac{4((2N)^{d_{vc}}+1)}{\delta}}$.
Thus $N \geq \frac{8}{\epsilon^2}\ln\left(\frac{4((2N)^{d_{vc}}+1)}{\delta}\right)$.
Here, $d_{vc} = 10$, $\epsilon = 0.05$, $\delta = 0.05$, so:

$$N \geq \frac{8}{0.05^2}\ln\left(\frac{4((2N)^{10}+1)}{0.05}\right)$$
$$N \approx 452957$$

Thus the sample size N needs to be greater than 452957.