CSCI 4100 Assignment 5

Boliang Yang 661541863

1. Exercise 2.8 in LFD
   (a) By definition, $\bar{g}(x) = \frac{1}{K}\sum_{k=1}^{K} g_k(x)$, then $\bar{g}(x)$ is a linear combination of $g_k(x) \in H$. Since H is closed under linear combination, then $\bar{g} \in H$.
   (b) The simple model is a binary classification. If the model's hypothesis set H contains only two hypotheses: one that always returns +1 and one that always returns -1. Then for dataset that generated +1 and -1 with equal probability, $\bar{g}(x)$ is 0 which clearly will not be in H.
   (c) No, $\bar{g}(x)$ will most likely to be somewhere between -1 and +1, and thus cannot be a binary function.

2. Problem 2.14 in LFD
   (a) From problem we know, $H_1, H_2, \dots H_K$ each has break point $d_{vc} + 1$, then $m_{H_i}(d_{vc} + 1) < 2^{d_{vc}+1}$ for all H_i where $i = 1, \dots k$. Then for $H = H_1 \cup H_2 \cup \dots \cup H_K$, $m_H(K(d_{vc} + 1)) \le \prod_{i=1}^{K} m_{H_i}(d_{vc} + 1) < (2^{d_{vc}+1})^K = 2^{K(d_{vc}+1)}$. Since $m_H(K(d_{vc} + 1)) < 2^{K(d_{vc}+1)}$, we have $d_{vc}(H) < K(d_{vc} + 1)$.
   (b) Since $m_{H_i}(l) \le l^{d_{vc}} + 1$ for all H_i where $i = 1, \dots k$, then $m_H(l) \le \sum_{i=1}^{K} m_{H_i}(l) \le K(l^{d_{vc}} + 1) = Kl^{d_{vc}} + K$. Since $2^l > 2Kl^{d_{vc}}$, then $2^l > Kl^{d_{vc}} + Kl^{d_{vc}} \ge Kl^{d_{vc}} + K$. We have $m_H(l) < 2^l$, and it follows that $d_{vc}(H) \le l$.
   (c) Assume l be the value $7(d_{vc} + K)\log_2(d_{vc}K)$, and we get:

$$2^l > 2Kl^{d_{vc}}$$

$$2^{7(d_{vc}+K)\log_2(d_{vc}K)} > 2K[7(d_{vc} + K)\log_2(d_{vc}K)]^{d_{vc}}$$

$$(d_{vc}K)^{7(d_{vc}+K)} > 2\,K[7(d_{vc} + K)\log_2(d_{vc}K)]^{d_{vc}}$$

   The inequality holds since the left-hand-side grows much faster than the right-hand-side: exponent in the left-hand-side is $7(d_{vc} + K)$, which is much greater than exponent d_{vc} in the right-hand-side. Then $l = 7(d_{vc} + K)\log_2(d_{vc}K)$ satisfies $2^l > 2Kl^{d_{vc}}$. From (a)&(b), $d_{vc}(H) < K(d_{vc} + 1)$, $d_{vc}(H) \le l$ where $2^l > 2Kl^{d_{vc}}$. Thus, $d_{vc}(H) \le \min(K(d_{vc} + 1), 7(d_{vc} + K)\log_2(d_{vc}K))$.

3. Problem 2.15 in LFD

(a) One example of monotonic classifier in two dimensions is $h(x, y) = sign(x + y)$: when $x + y \geq 0$ returns +1, when $x + y < 0$ returns -1. Since for any two points (x_1, y_1), (x_2, y_2), if $x_1 \leq x_2,\ y_1 \leq y_2$, we have $sign(x_1 + y_1) \leq sign(x_2 + y_2)$, then the classifier is monotonic.

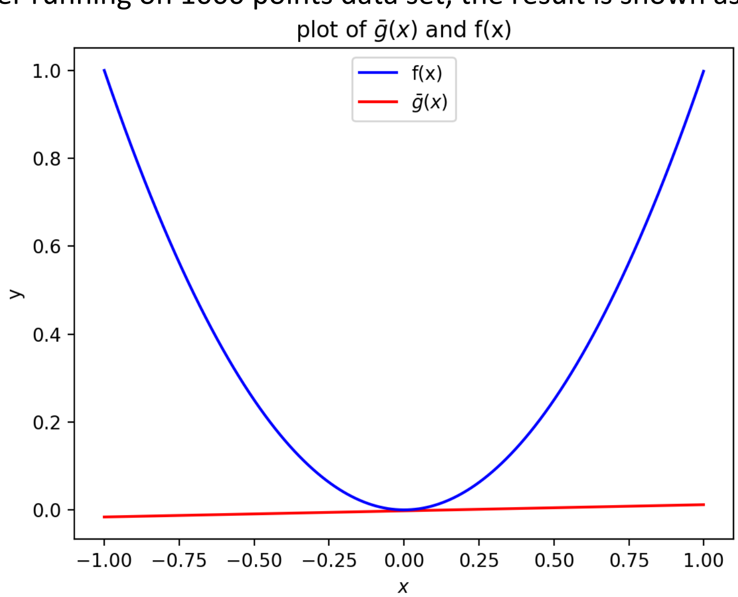
(b) We first choose one point, and then generate the next point by increasing the first component (x) and decreasing the second component (y) in order to keep x+y unchanged until N points are obtained. In this case, all the points can always be shattered and all dichotomies exist. Then $m_H(N) = 2^N,\ d_{vc}(H) = \infty$.

4. Problem 2.24 in LFD

(a) Since $D = \{(x_1, x_1^2), (x_2, x_2^2)\}$, and hypothesis is of the form $h(x) = ax + b$, then we have slope $(x_1 + x_2)$, and intercept $-x_1x_2$. Since input variable x is uniformly distributed in $[-1, 1]$, then expected value $E(x_1) = E(x_2) = 0$. Thus, $\bar{g}(x) = 0$.

(b) First generate 1000 points data set with target function $f(x) = x^2$ where $x \in [-1,1]$. Then compute hypothesis $g_k(x)$ based on each point. Now we can compute the average function $\bar{g}(x)$ using $\bar{g}(x) = \frac{1}{1000}\sum_{k=1}^{1000} g_k(x)$. We can compute bias as $bias = \left(\bar{g}(x) - f(x)\right)^2$, compute var as $var = \mathbb{E}_D\left[\left(g^{(D)}(x) - \bar{g}(x)\right)^2\right]$, compute $\mathbb{E}[E_{out}]$ as $\mathbb{E}[E_{out}] = bias + var$.

(c) After running on 1000 points data set, the result is shown as following:



$\bar{g}(x) = 0.014003x - 0.002224$, $bias = 0.199232$, $var = 0.334724$, $\mathbb{E}[E_{out}] = 0.533956$

(d) Since $\bar{g}(x) = 0$, then $bias = \frac{1}{2}\int_{-1}^{1}\left(f(x) - \bar{g}(x)\right)^2 dx = \frac{1}{2}\int_{-1}^{1}(x^2 - 0)^2 dx = \frac{1}{5}$.

$var = \mathbb{E}_D\left[\left(g^{(D)}(x) - \bar{g}(x)\right)^2\right] = \mathbb{E}_D\left[\left(g^{(D)}(x)\right)^2\right] = \frac{1}{2}\int_{-1}^{1} x^2 dx = \frac{1}{3}$.

$\mathbb{E}[E_{out}] = bias + var = \frac{1}{5} + \frac{1}{3} = \frac{8}{15}$.