HW 5 – Idan Michaeli 208429878, Itamar Yacoby 209794403

1. a. *Let* $x, y \in R^n$, *then*

$$K(x,y) = 5K_1(x,y) + 4K_2(x,y) = 5\varphi_1(x)^T\varphi_1(y) + 4\varphi_2(x)^T\varphi_2(y)$$

We define $\varphi(x) = \begin{pmatrix} \sqrt{5}\varphi_1(x) \\ 2\varphi_2(x) \end{pmatrix}$ *therefore we get that*

$$5\varphi_1(x)^T\varphi_1(y) + 4\varphi_2(x)^T\varphi_2(y) = \left(\sqrt{5}\varphi_1(x)^T, 2\varphi_2(x)^T\right)\begin{pmatrix} \sqrt{5}\varphi_1(y) \\ 2\varphi_2(y) \end{pmatrix} = \varphi(x)^T\varphi(y)$$

Thus K is a kernel by definition and it's corresponding φ *is defined as above.*

b. *Let* K_1 *be a kernel and let* $\varphi_1(x): R^n \to R^m$ *s.t* $n < m$ *be its corresponding mapping and the data is linearly separable in the space* R^m. *let* $w = (w_1, \dots, w_m)$ *be the weight vector of the linear classifier. denote* $F(x) = w_0 + w_1\varphi_1^1 + \cdots + w_m\varphi_m^1$ *as the LC.*

Consider $K = 5K_1 + 4K_2$ *and* $\varphi(x) = \begin{pmatrix} \sqrt{5}\varphi_1(x) \\ 2\varphi_2(x) \end{pmatrix}$ *as found in last item, consider also that* $\varphi_2(x): R^s \to R^t$ *where* $s < t$.

let $w' = (w'_1, \dots, w'_m, \dots, w'_{m+t})$ *be a weight vector. We will show that* $G(x) = w'\varphi(x)$ *is a linear classifier:*

$$G(x) = w'_0 + w'_1\varphi_1 + \cdots + w'_{m+t}\varphi_{m+t} =$$
$$= w'_0 + w'_1\sqrt{5}\varphi_1^1 + \cdots + w'_m\sqrt{5}\varphi_m^1 + w'_{m+1}2\varphi_{m+1}^2 + \cdots + w'_{m+t}2\varphi_{m+t}^2$$

We set $w' = \left(w_0, \frac{w_1}{\sqrt{5}}, \dots, \frac{w_m}{\sqrt{5}}, \underbrace{0, \dots, 0}_{t}\right)$, *Therefore we get*

$$G(x) = w_0 + w_1\varphi_1^1 + \cdots + w_m\varphi_m^1 + 0 \cdot \varphi_{m+1}^2 + \cdots + 0 \cdot \varphi_{m+t}^2 =$$
$$= \boldsymbol{w_0 + w_1\varphi_1^1 + \cdots + w_m\varphi_m^1 = F(x)} \quad \textit{Which we know is a linear clasifier.}$$

Thus $w' = \left(w_0, \frac{w_1}{\sqrt{5}}, \dots, \frac{w_m}{\sqrt{5}}, \underbrace{0, \dots, 0}_{t}\right)$ *is the requested weight vector.*

c. *We define the following mapping* $\varphi: R \to R^N$ *as follows:*

$$\varphi(x) = \left(\underbrace{3,3, \dots 3,}_{x}\underbrace{0,0, \dots 0}_{N-x}\right)$$

W.L.O.G, assume that $x \le y$, *therefore we get* $\varphi(x)^T\varphi(y) = \begin{pmatrix} 3 \\ \vdots \\ 3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\left(\underbrace{3,3, \dots 3,}_{y}\underbrace{0,0, \dots 0}_{N-y}\right) =$

$$= \underbrace{3 \cdot 3 + \cdots + 3 \cdot 3}_{x} + \underbrace{3 \cdot 0 + \cdots + 3 \cdot 0}_{y-x} + \underbrace{0 \cdot 0 + \cdots + 0 \cdot 0}_{N-y} = 9(\underbrace{1 + \cdots + 1}_{x}) = 9x = 9\min(x,y)$$

$2. \textit{Consider } R(h,s) = 200 \cdot h^{\frac{2}{3}} \cdot s^{\frac{1}{3}}, \ \ \textit{Consider also } g(h,s) = 20h + 170s - 20{,}000$

$$L(h,s) = 200 \cdot h^{\frac{2}{3}} \cdot s^{\frac{1}{3}} + \lambda \cdot (20h + 170s - 20{,}000)$$

$$\frac{\partial}{\partial h} L(h,s) = \frac{400 \cdot s^{\frac{1}{3}}}{3 \cdot h^{\frac{1}{3}}} + 20\lambda = 0$$

$$\frac{\partial}{\partial s} L(h,s) = \frac{200 \cdot h^{\frac{2}{3}}}{3 \cdot s^{\frac{2}{3}}} + 170\lambda = 0$$

$$\frac{\partial}{\partial \lambda} L(h,s) = 20h + 170s - 20{,}000 = 0$$

$$\frac{\frac{400 \cdot s^{\frac{1}{3}}}{3 \cdot h^{\frac{1}{3}}}}{\frac{200 \cdot h^{\frac{2}{3}}}{3 \cdot s^{\frac{2}{3}}}} = \frac{-20\lambda}{-170\lambda} \quad \rightarrow \quad \frac{s}{h} = \frac{1}{17} \ \rightarrow \ h = 17 \cdot s$$

$$20 \cdot 17 \cdot s + 170s = 20{,}000$$

$$510 \cdot s = 20{,}000$$

$$s = \frac{20000}{510} = 39\frac{11}{51}$$

$$h = 17 \cdot 39\frac{11}{51} = 666\frac{2}{3}$$

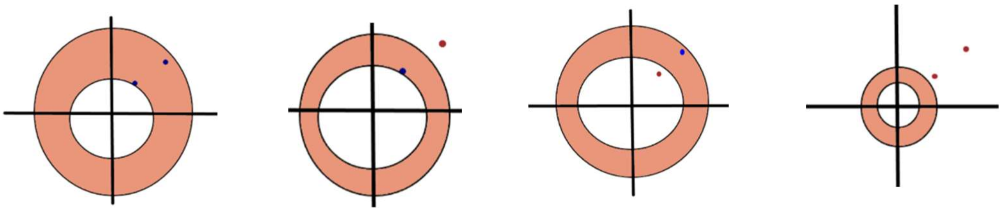
$$\lambda = \frac{200 \cdot \left(666\frac{2}{3}\right)^{\frac{2}{3}}}{-170 \cdot 3 \cdot \left(39\frac{11}{51}\right)^{\frac{2}{3}}} = -2.5927$$

Maximum Revenue:

$$200 \cdot \left(666\frac{2}{3}\right)^{\frac{2}{3}} \cdot \left(39\frac{11}{51}\right)^{\frac{1}{3}} = \mathbf{51,854.8158}\$$$

$3.a.$

$VC(H) \geq 2, we\ show\ that\ there\ exists\ 2\ points\ that\ we\ can\ shatter:$



$We\ prove\ now\ that\ VC(H) < 3:$
$Consider\ a\ set\ of\ 3\ distinct\ points\ \{v_1, v_2, v_3\}.$
$Consider\ a\ ring\ that\ contains\ the\ 2\ points\ with\ minimum\ and\ maximum\ euclidean\ distance.$
$There\ are\ at\ most\ 2\ such\ points, denote\ that\ set\ of\ 2\ points\ as\ S.$
$note\ that\ S \subset \{v_1, v_2, v_3\}.$
$Any\ ring\ that\ contains\ S, must\ also\ contain\ \{v_1, v_2, v_3\}, but\ there\ is\ at\ least\ one\ v_i\ that$
$was\ not\ used\ in\ S, but\ still\ must\ be\ in\ the\ ring.$
$Therefore, the\ label\ assignment\ that\ labels\ all\ the\ points\ in\ S\ with + and\ v_i\ with -$
$cannot\ be\ consistent\ with\ any\ ring.$

$\Rightarrow VC(H) = 2$

$b.$

$Let\ r_1^*, r_2^*\ s.t\ r_1^*, \leq r_2^*\ be\ the\ radiuses\ or\ the\ targeted\ ring.$
$Each\ instance\ viewed\ in\ the\ sample\ is\ drawn\ from\ an\ unknown\ distribution\ D\ and\ comprises$
$of\ 2\ features\ \big(position\ of\ the\ instance, (x1\ , x2\ )\big) and\ a\ target\ value, +1\ if\ it's\ inside\ the$
$ring\ and - 1\ otherwise.$
$The\ Algorithm:$

$We\ take\ the\ tightest\ ring\ so\ that\ all\ the\ positive\ points\ lie\ inide\ that\ ring.$
$(The\ maximum\ r_1\ and\ minimum\ r_2\ \ s.t\ \ every\ positive\ point\ (x_1, x_2)\ satisfies$
$\mathrm{x}_1^2 + x_2^2 \geq r_1\ \ and\ \ \mathrm{x}_1^2 + x_2^2 \leq r_2).$

$The\ algorithm\ outputs\ r_1, r_2\ \ s.t\ \ r_1^* \leq r_1\ \ and\ \ r_2 \leq r_2^*$
$We\ define\ \ r_\epsilon^1, r_\epsilon^2 = the\ \ radiuses\ s.t\ the\ sum\ of\ the\ error\ of\ the\ annuluses$
$A_{r_1} = h(r_1^*, r_1)\ and\ A_{r_2} = h(r_2, r_2^*)\ is\ smaller\ than\ \epsilon.$
$r_\epsilon^1, r_\epsilon^2 = argmax_{r_1} arginf_{r_2}\big(P_{r_1}\big[(x,y) \in A_{r_1}\big] + P_{r_2}\big[(x,y) \in A_{r_2}\big] \leq \epsilon\big).$

$There\ are\ 2\ cases:$
$case\ 1: r_1 \leq r_\epsilon^1\ \ and\ r_\epsilon^2 \leq r_2\ \ then\ the\ probability\ of\ both\ annuluses\ is\ smaller\ than\ \epsilon.$
$case\ 2: (r_1 < r_\epsilon^1\ \ and\ r_\epsilon^2 \leq r_2)\ \ or\ \ (r_1 \leq r_\epsilon^1\ \ and\ r_\epsilon^2 > r_2)\ or\ \ (r_1 > r_\epsilon^1\ \ and\ r_\epsilon^2 > r_2)\ then\ the$
$probability\ of\ missing\ the\ annulus\ h(r_1^*, r_\epsilon^1)\ or\ h(r_\epsilon^2, r_2^*)\ or\ both\ respectively\ is:$

$$p \leq 2 \cdot \left(1 - \frac{\epsilon}{2}\right)^m \leq 2 \cdot e^{-\frac{\epsilon m}{2}} \underbrace{\leq}_{m \geq \frac{2 \cdot \ln\left(\frac{2}{\delta}\right)}{\epsilon}} 2 \cdot e^{-\ln\left(\frac{2}{\delta}\right)} = 2 \cdot e^{\ln\left(\frac{\delta}{2}\right)} = \delta$$

$Thus\ we\ receive\ that\ our\ concept\ is\ PAC - learnable$

$Sample\ and\ Time\ Complexity:\ m(\epsilon, \delta) \in O\left(\frac{2}{\epsilon}, \ln\left(\frac{2}{\delta}\right)\right)$

$c. \quad \epsilon = 0.05, \;\; \delta = 0.05$

1. $m \geq \frac{1}{0.05}\left(4\log\frac{2}{0.05} + 16\log\frac{13}{0.05}\right) \approx 2993$

2. $m \geq \frac{2 \cdot \ln\left(\frac{2}{0.05}\right)}{0.05} \approx 147$

$We\ got\ a\ smaller\ m\ in\ the\ bound\ we\ found\ in\ item\ b, because\ in\ the\ direct\ approach\ we\ use\ the\ characeristic\ of\ the\ specific\ distribution\ rather\ than\ the\ VC\ bound\ which\ needs\ to\ apply\ to\ every\ distribution\ with\ the\ same\ VC\ dimension.$

$4.a.\ VC(H_3) \geq 4,\ \ this\ is\ the\ proof:$
$Consider\ any\ set\ of\ 4\ distinct\ points\ \{v_1, v_2, v_3, v_4\}\ s.t\ \ v_i \in \mathbb{R}\ \forall 1 \leq i \leq 4$
$and\ \ v_1 < v_2 < v_3 < v_4.$
$given\ m = 3\ we\ get\ a\ 7\ node\ full\ binary\ tree\ with\ 4\ leaves\ where\ each\ can\ classify\ as\ +\ or\ -.$
$Every\ non-leaf\ node\ represent\ a\ number\ in\ \mathbb{R}\ \ s.t\ \ a\ point\ v_i\ percolates\ down\ to\ the\ left\ child\ if\ v_i\ smaller\ than\ the\ node\ threshold, and\ percolates\ down\ to\ the\ right\ child\ if\ v_i\ greater\ or\ equal\ than\ the\ node\ threshold.$
$Every\ leaf\ represents\ an\ interval\ that\ is\ distinct\ from\ the\ other\ children\ (from\ the\ construction\ of\ the\ tree).\ Consequently, each\ leaf\ can\ represent\ a\ label\ for\ all\ the\ points\ in\ this\ interval.$
$Thus, any\ dichotomy\ of\ the\ points\ can\ classified\ by\ these\ 4\ leaves.$
$We\ prove\ now\ that\ VC(H_3) < 5:$
$Consider\ a\ set\ of\ 5\ points\ \{v_1, v_2, v_3, v_4, v_5\},$
$v_5\ percolate\ down\ in\ the\ tree\ same\ as\ one\ of\ the\ v_i (\forall 1 \leq i \leq 4)$
$and\ therefore\ belong\ to\ the\ same\ interval\ as\ this\ v_i.$
$Suppose\ that\ v_i\ is\ classified\ as + and\ v_5\ as\ -, therefore\ we\ get\ that\ the\ label\ assignment\ is\ not\ consistent\ (all\ the\ points\ in\ the\ interval\ reprsented\ by\ the\ leaf\ suppose\ to\ be\ classified\ the\ same).$
$This\ means\ that\ there\ is\ no\ set\ of\ size\ 5\ that\ can\ be\ shattered\ by\ H,\ \ and\ therefore\ VC(H) < 5.$

$\Rightarrow VC(H) = 4.$

$b.$
$VC(H) = 2^{m-1}$ $\ which\ represents\ the\ number\ of\ leaves\ in\ a\ height\ m\ full\ binary\ tree.$
$Proof$:
$First\ we\ prove\ that\ VC(H) \geq 2^{m-1}$:
$Consider\ any\ set\ of\ 2^{m-1}\ distinct\ points\ \{v_1, v_2, \dots, v_{2^{m-1}}\}\ \ s.t\ \ v_i \in \mathbb{R}\ \ \forall 1 \leq i \leq 2^{m-1}$
$and\ \ v_1 < v_2 < \cdots < v_{2^{m-1}}.$
$given\ m\ we\ get\ a\ \ 2^m - 1\ \ node\ full\ binary\ tree\ with\ 2^{m-1}\ leaves.$
$Every\ non-leaf\ node\ represent\ a\ number\ in\ \mathbb{R}\ \ s.t\ \ a\ point\ v_i\ percolates\ down\ to\ the\ left$
$child\ if\ v_i\ smaller\ than\ the\ node\ threshold, and\ percolates\ down\ to\ the\ right\ child\ if$
$v_i\ greater\ or\ equal\ than\ the\ node\ threshold.$
$Every\ leaf\ represents\ an\ interval\ that\ is\ distinct\ from\ the\ other\ children\ (from\ the\ construction$
$of\ the\ tree).\ \ Consequently, each\ leaf\ can\ represent\ a\ label\ for\ all\ the\ points\ in\ this\ interval.$
$Thus, any\ dichotomy\ of\ the\ points\ can\ classified\ by\ these\ 2^{m-1}\ leaves.$
$We\ prove\ now\ that\ VC(H_3) < 2^{m-1} + 1$:
$Suppose\ we\ add\ another\ point, v'.$
$v'\ percolate\ down\ in\ the\ tree\ same\ as\ one\ of\ the\ v_i\ (\forall 1 \leq i \leq 2^{m-1}).$
$Suppose\ that\ v_i\ is\ classified\ as + and\ v'\ as -, therefore\ we\ get\ that\ the\ label\ assignment$
$is\ not\ consistent.$
$This\ means\ that\ there\ is\ no\ set\ of\ size\ 2^{m-1} + 1\ that\ can\ be\ shattered\ by\ H$
$and\ therefore\ VC(H) < 2^{m-1} + 1\ .$

$\Rightarrow VC(H) = 2^{m-1}.$