HOMEWORK 2

YUNLIN ZHANG

A. Problem A

A.1.

B. Problem B

B.1. $\{X^+ \cup \{x_{m+1}\}, X^-\}$ **and** $\{X^+, X^- \cup \{x_{m+1}\}\}$ **are both linearly separable dichotomies by hyperplanes going through the origin (A) iff** $\{X^+, X^-\}$ **is linearly separable by a hyperplane going through the origin and** x_{m+1} **(B).**

First, to show (A)$\Rightarrow$(B)
(A)$\Rightarrow \exists \mathbf{w}_1$ at origin $s.t.\ \mathbf{w}_1 \cdot \mathbf{x}^+ > 0\ \forall \mathbf{x}^+ \in X^+ \cup \{x_{m+1}\},\ \mathbf{w}_1 \cdot \mathbf{x}^- < 0\ \forall \mathbf{x}^- \in X^-$
and
$\exists \mathbf{w}_2$ at origin $s.t.\ \mathbf{w}_2 \cdot \mathbf{x}^+ > 0\ \forall \mathbf{x}^+ \in X^+,\ \mathbf{w}_2 \cdot \mathbf{x}^- < 0\ \forall \mathbf{x}^- \in X^- \cup \{x_{m+1}\}$
Define the hyperplane defined by $\mathbf{w}_1$ as U_1 and by $\mathbf{w}_2$ as U_2

Consider $f(\alpha) = [\alpha \mathbf{w}_1 + (1-\alpha)\mathbf{w}_2] \cdot \mathbf{x}_{m+1},\ \alpha \in [0,1]$
$f(\alpha)$ is a linear function of α therefore is continuous in $\alpha \in [0,1]$ and $f(0) < 0,\ f(1) > 0$, by intermediate value theorem $\exists \alpha' \in (0,1)\ s.t.\ f(\alpha') = 0$
$\Rightarrow \mathbf{y} = \alpha' \mathbf{w}_1 + (1-\alpha')\mathbf{w}_2\ s.t.\ \mathbf{y} \cdot \mathbf{x}_{m+1} = 0$
$\Rightarrow \mathbf{y}$ is a normal vector that defines a hyperplane U containing $\mathbf{x}_{m+1}$
$\because \alpha', (1-\alpha') > 0$
$\mathbf{y} \cdot \mathbf{x}^+ = \alpha' \mathbf{w}_1 \cdot \mathbf{x}^+ + (1-\alpha')\mathbf{w}_2 \cdot \mathbf{x}^+ > 0, \forall \mathbf{x}^+ \in X^+$
$\mathbf{y} \cdot \mathbf{x}^- = \alpha' \mathbf{w}_1 \cdot \mathbf{x}^- + (1-\alpha')\mathbf{w}_2 \cdot \mathbf{x}^- < 0, \forall \mathbf{x}^- \in X^-$
$\Rightarrow \{X^+, X^-\}$ is linearly separable by the hyperplane defined by $\mathbf{y}$

By construction, since $\alpha' \in (0,1)$, U must be between U_1 and U_2
By squeeze theorem U_1 and U_2 go through the origin$\Rightarrow U$ also goes through the origin
$\Rightarrow \{X^+, X^-\}$ is a dichotomy that is linearly separable by a hyperplane that goes through the origin and $\mathbf{x}_{m+1}$. This completes the proof for (A)$\Rightarrow$(B)

Now, to show (B)$\Rightarrow$(A)
Consider the family of planes defined by the set of vectors $\{\mathbf{y}(\epsilon) = (1-\epsilon)\mathbf{w} + \epsilon \mathbf{x}_{m+1} : \epsilon \in (0,1]\}$
Using this definition,
$\mathbf{y} \cdot \mathbf{x}_{m+1} = \epsilon ||\mathbf{x}_{m+1}||^2 > 0$
We'd want to find an ϵ where $\mathbf{y} \cdot \mathbf{x} > 0,\ \forall \mathbf{x} \in X^+$ and $\mathbf{y} \cdot \mathbf{x} < 0,\ \forall \mathbf{x} \in X^-$
Denote any point in X^+ as $\mathbf{x}^+$ and in X^- as $\mathbf{x}^-$:
For points in X^+ to be correctly classified,
$\mathbf{y} \cdot \mathbf{x}^+ = (1-\epsilon)\mathbf{w} \cdot \mathbf{x}^+ + \epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}^+ > 0$
$\Leftrightarrow (1-\epsilon)\mathbf{w} \cdot \mathbf{x}^+ > |\epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}^+|$
$\Leftrightarrow (1-\epsilon)/\epsilon > |\mathbf{x}_{m+1} \cdot \mathbf{x}^+|/\mathbf{w} \cdot \mathbf{x}^+$ $\quad \because \epsilon > 0,\ \mathbf{w} \cdot \mathbf{x}^+ > 0$
$\Leftrightarrow 1/\epsilon > |\mathbf{x}_{m+1} \cdot \mathbf{x}^+|/\mathbf{w} \cdot \mathbf{x}^+ + 1$
$\Leftrightarrow \epsilon < [|\mathbf{x}_{m+1} \cdot \mathbf{x}^+|/\mathbf{w} \cdot \mathbf{x}^+ + 1]^{-1} = \delta_1 > 0$
Choose $\delta_1' = \min_{\mathbf{x}^+} \delta_1$
$\therefore$ if $\epsilon < \delta_1'$ then all the points in X^+ will still be correctly classified
Similarly, for points in X^- to be correctly classified,

$\mathbf{y} \cdot \mathbf{x}^- = (1-\epsilon)\mathbf{w} \cdot \mathbf{x}^- + \epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}^- < 0$
$\Leftrightarrow (1-\epsilon)\mathbf{w} \cdot \mathbf{x}^- < -|\epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}|$
$\Leftrightarrow (1-\epsilon)/\epsilon > -|\epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}|/\mathbf{w} \cdot \mathbf{x}^-$ $\quad \because \mathbf{w} \cdot \mathbf{x}^- < 0$
$\Leftrightarrow \epsilon < [-|\mathbf{x}_{m+1} \cdot \mathbf{x}^+|/\mathbf{w} \cdot \mathbf{x}^- + 1]^{-1} = \delta_2 > 0$
Choose $\delta_2' = \min_{\mathbf{x}^-} \delta_2$
$\therefore$ if $\epsilon < \delta_2'$ then all the points in X^- will still be correctly classified
So if we choose some $\epsilon' < \min(\delta_1', \delta_2') \Rightarrow$
$\mathbf{y}(\epsilon') \cdot \mathbf{x} > 0 \ \forall \mathbf{x} \in X^+$
$\mathbf{y}(\epsilon') \cdot \mathbf{x} < 0 \ \forall \mathbf{x} \in X^-$
$\mathbf{y}(\epsilon') \cdot \mathbf{x}_{m+1} > 0$
Therefore the dichotomy $\{X^+ \cup \{\mathbf{x}_{m+1}\}, X^-\}$ is linearly separable by $\mathbf{y}(\epsilon')$

Similar, to show that the dichotomy $\{X^+, X^- \cup \{\mathbf{x}_{m+1}\}$ is linearly separable, we use the family of planes defined by vectors in $\{\mathbf{z}(\epsilon) = (1-\epsilon)\mathbf{w} - \epsilon \mathbf{x}_{m+1} \ \epsilon \in (0,1]\}$
$\Rightarrow \mathbf{z} \cdot \mathbf{x}_{m+1} = -\epsilon ||\mathbf{x}_{m+1}||^2 < 0$
The conditions:
$\mathbf{z} \cdot \mathbf{x}^+ = (1-\epsilon)\mathbf{w} \cdot \mathbf{x}^+ - \epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}^+ > 0$
and
$\mathbf{z} \cdot \mathbf{x}^- = (1-\epsilon)\mathbf{w} \cdot \mathbf{x}^- - \epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}^- < 0$
simplify to the same forms:
$(1-\epsilon)\mathbf{w} \cdot \mathbf{x}^+ > |\epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}^+|$
and
$(1-\epsilon)\mathbf{w} \cdot \mathbf{x}^- < -|\epsilon \mathbf{x}_{m+1} \cdot \mathbf{x}|$
So we can choose the same ϵ' as defined in the previous part and the following conditions will hold:
$\mathbf{z}(\epsilon') \cdot \mathbf{x} > 0 \ \forall \mathbf{x} \in X^+$
$\mathbf{z}(\epsilon') \cdot \mathbf{x} < 0 \ \forall \mathbf{x} \in X^-$
$\mathbf{z}(\epsilon') \cdot \mathbf{x}_{m+1} < 0$
Therefore the dichotomy $\{X^+ \cup \{\mathbf{x}_{m+1}\}, X^-\}$ is linearly separable by $\mathbf{z}(\epsilon')$, and this completes the proof for (B)$\Rightarrow$(A) $\quad \square$

B.2. Let X=$\{\mathbf{x}_1, ..., \mathbf{x}_m\} \subset \mathbb{R}^d$ *s.t.* any d-element subset of X is linearly independent. Then the number of linearly separable labelings of X is $C(m,d) = 2\sum_{k=0}^{d-1} \binom{m-1}{k}$.
Following the proof for Sauer's Lemma

C. Problem C

D. Problem D

D.1. Show $K(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{N} \cos^n(x_i^2 - y_i^2) \quad \forall(\mathbf{x},\mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N$ is PDS.

Using $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} \cdot \begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix}$

Let $\Phi_i(\mathbf{x}) = \begin{bmatrix} \cos x_i^2 \\ \sin x_i^2 \end{bmatrix}$ and $K_i(\mathbf{x},\mathbf{y}) = \Phi_i(\mathbf{x}) \cdot \Phi_i(\mathbf{y})$ is PDS

$K(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{N} [K_i(\mathbf{x},\mathbf{y})]^n \Rightarrow K(\mathbf{x},\mathbf{y})$ is PDS by closure of PDS kernels. $\square$

D.2. Show $K(\mathbf{x},\mathbf{y}) = \exp(-||\mathbf{x}-\mathbf{y}||/\sigma) \quad \forall(\mathbf{x},\mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N$ is PDS.

Consider if $\sum_{i=1}^{m} c_i = 0$,

$\sum_{i,j=1}^{m} c_i c_j ||\mathbf{x}_i - \mathbf{x}_j|| = \sum_{i,j}^{m} c_i c_j \sqrt{(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j)}$

Case 1: $(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j) \le 1$

$\sum_{i,j=1}^{m} c_i c_j ||\mathbf{x}_i - \mathbf{x}_j|| = \sum_{i,j} c_i c_j \sqrt{(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j)}$

$\quad \le \sum_{i,j} c_i c_j = \sum_i c_i \sum_j c_j = 0$

Case 2: $(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j) > 1 \Rightarrow \sqrt{(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j)} \le (\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j)$

$\sum_{i,j=1}^{m} c_i c_j ||\mathbf{x}_i - \mathbf{x}_j|| = \sum_{i,j} c_i c_j \sqrt{(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j)}$

$\quad \le \sum_{i,j} c_i c_j (\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{x}_i - \mathbf{x}_j)$

$\quad = \sum_{i,j} c_i c_j (||\mathbf{x}_i||^2 + ||\mathbf{x}_j||^2 - 2\mathbf{x}_i \cdot \mathbf{x}_j)$

$\quad = \sum_{i,j} c_i c_j (||\mathbf{x}_i||^2 + ||\mathbf{x}_j||^2) - 2\sum_{i,j} c_i c_j (\mathbf{x}_i \cdot \mathbf{x}_j)$

$\quad = \sum_{i,j} c_i c_j (||\mathbf{x}_i||^2 + ||\mathbf{x}_j||^2) - 2\sum_i c_i \mathbf{x}_i \cdot \sum_j c_j \mathbf{x}_j$

$\quad = \sum_{i,j} c_i c_j (||\mathbf{x}_i||^2 + ||\mathbf{x}_j||^2) - 2||\sum_i c_i \mathbf{x}_i||^2$

$\quad \le \sum_{i,j} c_i c_j (||\mathbf{x}_i||^2 + ||\mathbf{x}_j||^2)$

$\quad = \sum_i c_i \sum_j c_j ||\mathbf{x}_j||^2 + \sum_j c_j \sum_i c_i ||\mathbf{x}_i||^2 = 0 \qquad \because \sum_i c_i = 0$

$\Rightarrow \sum_{i,j=1}^{m} c_i c_j ||\mathbf{x}_i - \mathbf{x}_j|| \le 0, \ \forall(\mathbf{x}_i, \mathbf{x}_j) \in \mathbb{R}^N \times \mathbb{R}^N$

$\Rightarrow \ K_0(\mathbf{x},\mathbf{y}) = ||\mathbf{x}-\mathbf{y}||$ is an NDS kernel as defined in textbook.

$\Rightarrow \ K(\mathbf{x},\mathbf{y}) = \exp(-K_0/\sigma)$, and $1/\sigma > 0 \Rightarrow K(\mathbf{x},\mathbf{y})$ is PDS. $\square$