Stav Yosef, 316298876
Stav Vaknin, 313581654

Machine Learning from Data - Homework 5

Question 1

1.a

We know that K, L are kernels, therefore exists two mappings φ_1, φ_2 that
$\mathrm{K}(\mathrm{x}, \mathrm{y}) = \langle \varphi_1(\mathrm{x}), \varphi_1(\mathrm{y}) \rangle,\ \alpha \cdot \mathrm{K}(\mathrm{x}, \mathrm{y}) = \langle \sqrt{\alpha} \cdot \varphi_1(\mathrm{x}), \sqrt{\alpha} \cdot \varphi_1(\mathrm{y}) \rangle$
and
$\mathrm{L}(\mathrm{x}, \mathrm{y}) = \langle \varphi_2(\mathrm{x}), \varphi_2(\mathrm{y}) \rangle,\ \beta \cdot \mathrm{L}(\mathrm{x}, \mathrm{y}) = \langle \sqrt{\beta} \cdot \varphi_2(\mathrm{x}), \sqrt{\beta} \cdot \varphi_2(\mathrm{y}) \rangle$

Define $\mathrm{R}(\mathrm{x}, \mathrm{y}) = \alpha \cdot \mathrm{K}(\mathrm{x}, \mathrm{y}) + \beta \cdot \mathrm{L}(\mathrm{x}, \mathrm{y})$ for all $\alpha, \beta > 0$

$$\mathrm{R}(\mathrm{x}, \mathrm{y}) = \alpha \cdot \mathrm{K}(\mathrm{x}, \mathrm{y}) + \beta \cdot \mathrm{L}(\mathrm{x}, \mathrm{y})$$
$$= \langle \sqrt{\alpha} \cdot \varphi_1(\mathrm{x}), \sqrt{\alpha} \cdot \varphi_1(\mathrm{y}) \rangle + \langle \sqrt{\beta} \cdot \varphi_2(\mathrm{x}), \sqrt{\beta} \cdot \varphi_2(\mathrm{y}) \rangle$$

We know that inner product space is vector space, therefore we know that
$(\mathrm{x}_1, \mathrm{y}_1) + (\mathrm{x}_2, \mathrm{y}_2) = (\mathrm{x}_1 + \mathrm{x}_2, \mathrm{y}_1 + \mathrm{y}_2)$ and in our case,

$$= \langle \sqrt{\alpha} \cdot \varphi_1(\mathrm{x}) + \sqrt{\beta} \cdot \varphi_2(\mathrm{x}), \sqrt{\alpha} \cdot \varphi_1(\mathrm{y}) + \sqrt{\beta} \cdot \varphi_2(\mathrm{y}) \rangle$$

$\mathrm{R}(\mathrm{x}, \mathrm{y})$ is kernel because we expressed as inner product of K, L mappers, hence R is kernel.

1.b

i. Let $\mathrm{K}(\mathrm{x}, \mathrm{y}) = 2\mathrm{x}^{\mathrm{T}}\mathrm{y}, \mathrm{L}(\mathrm{x}, \mathrm{y}) = \mathrm{x}^{\mathrm{T}}\mathrm{y}$ be non-zero kernels
$(\mathrm{K} - \mathrm{L})(\mathrm{x}, \mathrm{y}) = \mathrm{x}^{\mathrm{T}}\mathrm{y}$ which is polynomial kernel → $\mathrm{K} - \mathrm{L}$ is a kernel.

ii. Let $\mathrm{K}(\mathrm{x}, \mathrm{y}) = \mathrm{x}^{\mathrm{T}}\mathrm{y}, \mathrm{L}(\mathrm{x}, \mathrm{y}) = 2\mathrm{x}^{\mathrm{T}}\mathrm{y}$ be non-zero kernels
$(\mathrm{K} - \mathrm{L})(\mathrm{x}, \mathrm{y}) = -(\mathrm{x}^{\mathrm{T}}\mathrm{y})$ is not a kernel. (we know K is non-zero kernel so $\mathrm{K}(\mathrm{x}, \mathrm{y}) > 0$ and if $\left((\mathrm{x}, \mathrm{y}) \mapsto -(\mathrm{x}^{\mathrm{T}}\mathrm{y})\right)$ is a kernel its need to be bigger than 0 → contradiction.)

Stav Yosef, 316298876
Stav Vaknin, 313581654

Question 2

$f(x,y,z) = x^2 + y^2 + z^2, g(x,y,z) = \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} + \frac{z^2}{\beta^2} - 1 = 0$
$\mathcal{L}(x,y,z,\lambda) = f(x,y,z) + \lambda \cdot g(x,y,z)$

Calculate $\mathcal{L}(x,y,z,\lambda)$ partial derivative and solve $\mathcal{L}(x,y,z,\lambda) = 0$

$$\begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix} = -\lambda \begin{bmatrix} \frac{2x}{\alpha^2} \\ \frac{2y}{\beta^2} \\ \frac{2z}{\beta} \end{bmatrix} \rightarrow \begin{cases} 2x + \lambda \cdot \frac{2x}{\alpha^2} = 0 \rightarrow x = 0 \text{ or } \lambda = -\alpha^2 \\ 2y + \lambda \cdot \frac{2y}{\beta^2} = 0 \rightarrow y = 0 \text{ or } \lambda = -\beta^2 \\ 2z + \lambda \cdot \frac{2z}{\beta^2} = 0 \rightarrow z = 0 \text{ or } \lambda = -\beta^2 \end{cases} \rightarrow \text{we can infer that } y = z$$

We know that $\alpha > \beta > 0$, hence $x = 0$ or $y = z = 0$
If $x = 0$
$\frac{y^2}{\beta^2} + \frac{z^2}{\beta^2} = 1 \rightarrow y^2 + z^2 = \beta^2 \rightarrow 2y^2 = \beta^2 \rightarrow y = z = \frac{\beta}{\sqrt{2}}$
If $y = z = 0$
$\frac{x^2}{\alpha^2} = 1 \rightarrow x = \alpha$

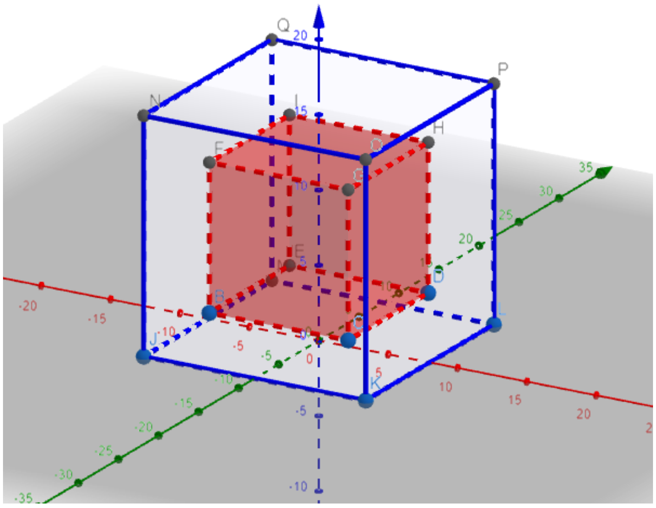
$f\left(x = 0, y = z = \frac{\beta}{\sqrt{2}}\right) = \beta^2$
$f(x = \alpha, y = z = 0) = \alpha^2$

We know that $\alpha > \beta > 0$, therefore
$f(x = \alpha, y = z = 0)$ is **maximum** point.
$f\left(x = 0, y = z = \frac{\beta}{\sqrt{2}}\right)$ is **minimum** point.

Stav Yosef, 316298876
Stav Vaknin, 313581654

Question 3

We know that our samples come from $\mathbb{R}^3$ and $\mathrm{C} = \mathrm{H}$ which is the origin centered boxes. We need to find the hypothesis which separates the samples into two groups, group 1 (the red box) and group 0 (the blue box).



We have 3 parameters $\mathrm{a}, \mathrm{b}, \mathrm{c} \in \mathbb{R}^3$ and we need to bound the box using these parameters. Our algorithm will go through randomized instances (m samples) and keep the maximum positive distance from the origin (for x, y, z separately) then we can say that:

$\mathrm{a} = \mathrm{x}_{\max} = |\mathrm{x}_{\min}| = |-\mathrm{a}|$
$\mathrm{b} = \mathrm{y}_{\max} = |\mathrm{y}_{\min}| = |-\mathrm{b}|$
$\mathrm{c} = \mathrm{z}_{\max} = |\mathrm{z}_{\min}| = |-\mathrm{c}|$
In the final step we can connect edges between correct points

1. $(-\mathrm{a}, -\mathrm{b}, -\mathrm{c})$ to $(\mathrm{a}, -\mathrm{b}, -\mathrm{c})$
2. $(-\mathrm{a}, \mathrm{b}, -\mathrm{c})$ to $(\mathrm{a}, \mathrm{b}, -\mathrm{c})$
3. $(-\mathrm{a}, -\mathrm{b}, \mathrm{c})$ to $(\mathrm{a}, -\mathrm{b}, \mathrm{c})$
4. $(-\mathrm{a}, \mathrm{b}, \mathrm{c})$ to $(\mathrm{a}, \mathrm{b}, \mathrm{c})$
5. $(-\mathrm{a}, -\mathrm{b}, -\mathrm{c})$ to $(-\mathrm{a}, \mathrm{b}, -\mathrm{c})$
6. $(\mathrm{a}, -\mathrm{b}, -\mathrm{c})$ to $(\mathrm{a}, \mathrm{b}, -\mathrm{c})$
7. $(-\mathrm{a}, -\mathrm{b}, \mathrm{c})$ to $(-\mathrm{a}, \mathrm{b}, \mathrm{c})$
8. $(\mathrm{a}, -\mathrm{b}, \mathrm{c})$ to $(\mathrm{a}, \mathrm{b}, \mathrm{c})$
9. $(-\mathrm{a}, -\mathrm{b}, -\mathrm{c})$ to $(-\mathrm{a}, -\mathrm{b}, \mathrm{c})$
10. $(\mathrm{a}, -\mathrm{b}, -\mathrm{c})$ to $(\mathrm{a}, -\mathrm{b}, \mathrm{c})$
11. $(-\mathrm{a}, \mathrm{b}, -\mathrm{c})$ to $(-\mathrm{a}, \mathrm{b}, \mathrm{c})$
12. $(\mathrm{a}, \mathrm{b}, -\mathrm{c})$ to $(\mathrm{a}, \mathrm{b}, \mathrm{c})$

Time complexity

Each space requires m samples and we iterated only 3 times $(\mathrm{a}, \mathrm{b}, \mathrm{c})$ then in total $\mathrm{O}(3\mathrm{m}) = \mathrm{O}(\mathrm{m})$

Stav Yosef, 316298876
Stav Vaknin, 313581654

Sample complexity

First, divide the place between the concept and the hypothesis (red & blue boxes) into 6 faces. We'll define the total volume of those "faces" as ϵ → each of the faces $= \frac{\epsilon}{6}$.
The probability of a point $\mathrm{D} \in \mathrm{X}^{\mathrm{m}}$ to be in either of those faces is $1 - \frac{\epsilon}{6}$
For a given ϵ and δ, the number of samples that needed is

$$\mathrm{P}(\{\mathrm{D} \in \mathrm{X}^{\mathrm{m}}: \mathrm{Err}(\mathrm{h} = \mathrm{L}(\mathrm{D}), \mathrm{concept}) > \epsilon\}) \leq 6\left(1 - \frac{\epsilon}{6}\right)^{\mathrm{m}} \leq 6\mathrm{e}^{-\frac{\mathrm{m}\epsilon}{6}}$$

$$6\mathrm{e}^{-\frac{\mathrm{m}\epsilon}{6}} \leq \delta$$

$$\ln\left(6\mathrm{e}^{-\frac{\mathrm{m}\epsilon}{6}}\right) \leq \ln(\delta)$$

$$\ln(6) + \ln\left(\mathrm{e}^{-\frac{\mathrm{m}\epsilon}{6}}\right) \leq \ln(\delta)$$

$$\ln(6) + \ln\left(-\frac{\mathrm{m}\epsilon}{6}\right) \leq \ln(\delta)$$

$$\ln(6) - \ln(\delta) \leq -\ln\left(-\frac{\mathrm{m}\epsilon}{6}\right)$$

$$\ln\left(\frac{6}{\delta}\right) \leq -\ln\left(-\frac{\mathrm{m}\epsilon}{6}\right)$$

$$\ln\left(\frac{6}{\delta}\right) \leq \frac{\mathrm{m}\epsilon}{6}$$

$$\frac{6}{\epsilon} \cdot \ln\left(\frac{6}{\delta}\right) \leq \mathrm{m}$$

Note - if we want confidence of $1 - \delta$ that we'll have error of ϵ we need at least $\frac{6}{\epsilon} \cdot \ln\left(\frac{6}{\delta}\right)$ instances.