ID: 802343 HW5

(1)

(a) Prove $\alpha K + \beta L$ is a kernel

Since K and L are kernels, there exist 2 mappings ψ_1 and ψ_2 such that

$$\forall x, x' : K(x, x') = \langle \psi_1(x), \psi_1(x') \rangle$$

$$\forall x, x' : L(x, x') = \langle \psi_2(x), \psi_2(x') \rangle$$

We will show that there is a mapping ψ_3, such that

$$\forall x, x' : (\alpha K + \beta L)(x, x') = \langle \psi_3(x), \psi_3(x') \rangle$$

Let $\psi_3(x) = (\sqrt{\alpha} \cdot \psi_1(x), \sqrt{\beta} \cdot \psi_2(x))$

Then, $\forall x, x' : \langle \psi_3(x), \psi_3(x') \rangle =$

$$= \langle (\sqrt{\alpha} \cdot \psi_1(x), \sqrt{\beta} \cdot \psi_2(x)), (\sqrt{\alpha} \cdot \psi_1(x'), \sqrt{\beta} \cdot \psi_2(x')) \rangle$$

$$= \langle (\sqrt{\alpha} \cdot \psi_1(x), \sqrt{\alpha} \cdot \psi_1(x') \rangle + \langle (\sqrt{\beta} \cdot \psi_2(x), \sqrt{\beta} \cdot \psi_2(x') \rangle$$

$$= \alpha \cdot \langle \psi_1(x), \psi_1(x') \rangle + \beta \cdot \langle \psi_2(x), \psi_2(x') \rangle$$

$$= \alpha \cdot K(x, x') + \beta \cdot L(x, x') = (\alpha K + \beta L)(x, x')$$

$\Rightarrow \alpha K + \beta L$ is a kernel

QED

(1) (b) (i) $K-L$ is a kernel

Let $L = \frac{1}{2}K$, then $K-L = K - \frac{1}{2}K = \frac{1}{2}K$

From the proof of (1a) we have that since K is a kenel then $\frac{1}{2}K$ is also a kernel

(ii) $K-L$ is not a kernel

Let $L = 2K$, then $K-L = K-2K = -K$

Assume by contradiction that $-K$ is a kernel, then there exists a mapping such that

$$\forall x, x' : -K(x, x') = \langle \varphi_1(x), \varphi(x') \rangle$$

Since K is a non-zero kernel, there exists an x, such that $K(x,x) > 0 \iff -K(x,x) < 0$, $-K(x,x) = \langle \varphi_1(x), \varphi_1(x^*) \rangle =$

$= \|\varphi_1(x)\| \geq 0$ which is a contradiction!

$\Rightarrow$ $-K$ is not a kernel

2

$$\begin{cases} 2x = \dfrac{2x\lambda}{\alpha^2} \\ 2y = \dfrac{2y\lambda}{\beta^2} \\ 2z = \dfrac{2z\lambda}{\beta^2} \end{cases} \Rightarrow \begin{cases} 2x - \dfrac{2x\lambda}{\alpha^2} = 0 \\ 2y - \dfrac{2y\lambda}{\beta^2} = 0 \\ 2z - \dfrac{2z\lambda}{\beta^2} = 0 \end{cases} \Rightarrow \begin{cases} 2x\left(1 - \dfrac{\lambda}{\alpha^2}\right) = 0 \\ 2y\left(1 - \dfrac{\lambda}{\beta^2}\right) = 0 \\ 2z\left(1 - \dfrac{\lambda}{\beta^2}\right) = 0 \end{cases}$$

$$g(x,y,z) = \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} + \frac{z^2}{\beta^2} = 1 \quad , \quad \alpha > \beta > 0$$

if $\lambda = 0$ then $\begin{cases} 2x = 0 \\ 2y = 0 \\ 2z = 0 \end{cases}$, so λ can't nullify all 3 equations

and we get these possible solutions:

$x=0, y=0 \Rightarrow z = \pm\beta$

$x=0, z=0 \Rightarrow y = \pm\beta$

$y=0, z=0 \Rightarrow x = \pm\alpha$

so the critical points are among the following points:

$(x,y,z) = (0,0,\pm\beta)$

$(x,y,z) = (0,\pm\beta,0)$

$(x,y,z) = (\pm\alpha,0,0)$

Since $\alpha > \beta > 0$,

$(\pm\alpha,0,0)$ - maximal points

and

$(0,0,\pm\beta), (0,\pm\beta,0)$ - minimal points

$$\Rightarrow \begin{aligned} \max(f) &= \alpha^2 \\ \min(f) &= \beta^2 \end{aligned}$$

③ $X = \mathbb{R}^3$.

$C = H = \{ h(a,b,c) = \{(x,y,z)$ s.t. $|x| \leq a, |y| \leq b, |z| \leq c\}$, s.t. $a,b,c \in \mathbb{R}_+ \}$

the set of all origin centered boxes.

The algorithm will produce a hypothesis which is the smallest relevant box that contains all the positive points. This can be done in $O(m)$ as follows (m - # of points):

Let $\Delta = \Delta^m = (x_i, y_i, z_i)_{i=1}^m$ be a set of points in $\mathbb{R}^3$, labeled positive and negative. Our algorithm seeks to return a hypothesis $h \in H$.

Let $(x_i, y_i, z_i)_{i=1}^{m^{(+)}}$ be all positive data points.

Find:

1) $l := \max_{1 \leq i \leq m^{(+)}} (|x_i|)$

2) $m := \max_{1 \leq i \leq m^{(+)}} (|y_i|)$

3) $n := \max_{1 \leq i \leq m^{(+)}} (|z_i|)$

l, m, n are the distances of the sides from the origin, ~~[illegible]~~ in the x, y, z directions respectively.

Consider $c \in C$ and let $\Delta^m(c) = (x_i(c), y_i(c), z_i(c))_{i=1}^m$ be the training data generated from c without errors and by drawing m independent points according to some probability distribution π on $\mathbb{R}^3$. We will denote the probability distribution thus induced on $(\mathbb{R}_3)^m$ by π^m.

(3) Given $\varepsilon > 0$ and $\delta > 0$ we compute $m(\varepsilon, \delta)$ so that:

$$(\text{eq. } 1) \qquad m \geq m(\varepsilon,\delta) \Rightarrow \underbrace{e(\Delta^m(c))}_{\text{r.v. that depends on } \Delta^m(c)} = \pi^m\big(\mathrm{err}_\pi(\underbrace{L(\Delta^m(c))}_{\text{hypothesis } h,\ \text{a box}}, c) > \varepsilon\big) \leq \delta$$

Now, consider the margins parallel to the sides of the box c. There are 6 of them and each 2 parallel margins are determined by one out of l, m, n parameters.

Let S_1, S_2, S_3 be the areas defined by both margins relevant to x, y, z axises respectively.

So that: $\quad \pi(S_1(\varepsilon)) = \pi(S_2(\varepsilon)) = \pi(S_3(\varepsilon)) = \varepsilon/3$

$$\Rightarrow \{\Delta^m(c) : \mathrm{err}_\pi(L(\Delta^m(c)), c) > \varepsilon\} \subseteq$$
$$\{\Delta^m(c) : \Delta^m(c) \cap S_1(\varepsilon) = \emptyset\} \cup$$
$$\{\Delta^m(c) : \Delta^m(c) \cap S_2(\varepsilon) = \emptyset\} \cup$$
$$\{\Delta^m(c) : \Delta^m(c) \cap S_3(\varepsilon) = \emptyset\}$$

because if $\Delta^m(c)$ visits the 3 strips then according to our construction the difference between c and $L(\Delta^m(c))$ will have

$$\pi \leq \pi(S_1(\varepsilon) \cup S_2(\varepsilon) \cup S_3(\varepsilon)) < \varepsilon$$

$$\Rightarrow \pi^m(\mathrm{err}_\pi(L(\Delta^m(c)), c) > \varepsilon) \leq$$
$$\pi^m(\Delta^m(c) \cap S_1(\varepsilon) = \emptyset) + \pi^m(\Delta^m(c) \cap S_2(\varepsilon) = \emptyset) + \pi^m(\Delta^m(c) \cap S_3(\varepsilon) = \emptyset) \leq$$
$$3\left(1 - \varepsilon/3\right)^m$$

Now, select $m(\varepsilon, \delta) = \frac{3}{\varepsilon}\left(\ln 3 + \ln\frac{1}{\delta}\right)$ so that eq1 holds.

QED