20.03.18

Homework 2
Supervised learning

1. from Mercer's theorem

$K(x,y)$ can be expressed:

$K(x,y) = \langle \varphi(x), \varphi(y) \rangle$ for some φ if $K(x,y)$ is positive semidefinite: (≥ 0)

$$\begin{bmatrix} K(x_1,x_2) & K(x_1,x_2) & \cdots \\ K(x_2,x_1) & & \\ & & \\ & & \end{bmatrix}$$

for any collection of $\{x_1 \ldots x_n\}$

I will use Gaussian kernel $\left(K(x,y) = e^{1/2\|x-y\|^2}\right) \geq 0$

define whether kernel is valid.

$X = \begin{bmatrix} a \\ b \end{bmatrix}$ $\quad X \cdot X$ $\begin{bmatrix} aa & ab \\ ab & bb \end{bmatrix}$ symmetry

a) $K(x,z) = K_1(x,z) + K_2(x,z)$

$K(x,z) = \langle \varphi_1(x), \varphi_1(z) \rangle + \langle \varphi_2(x), \varphi_2(z) \rangle$

$K(x,z) = \underbrace{e_1^{1/2\|x_1-z_1\|^2}}_{\geq 0} + \underbrace{e_2^{1/2\|x_2-z_2\|^2}}_{\geq 0}$

For any values in vectors x_1 and x_2 the sum is positive

So kernel is valid

b) $K(x,z) = K_1(x,z) - K_2(x,z) =$ $\quad x_1 \begin{bmatrix} a \\ b \end{bmatrix}$ $x_2 \begin{bmatrix} 2a \\ 2b \end{bmatrix}$ $\quad \underbrace{\begin{bmatrix} aa & ab \\ ab & bb \end{bmatrix} - \begin{bmatrix} 4aa & 4bb \\ 4bb & 4aa \end{bmatrix}}_{<0}$

$= e^{1/2\|x_1-z\|^2} - e^{1/2\|x_2-z\|}$. If choosing the same values for x_1 and x_2 vectors the difference $= 0$. Hence, this kernel is not valid.

$\boxed{\text{if } K_2 > K_1 \text{ difference could also be } < 0}$

c) $K(x,z) = a \cdot K_1(x,z) = a \cdot e^{1/2\|x-z\|^2}$ as $a > 0$ and $e^{1/2\|x-z\|^2} > 0$ the result > 0. $\quad X = \begin{bmatrix} a \\ b \end{bmatrix}$

So, kernel is valid

$$2 \begin{bmatrix} aa & ab \\ ab & bb \end{bmatrix} = \begin{bmatrix} 2aa & 2ab \\ 2ab & 2bb \end{bmatrix}$$

d) $K(x,z) = -a \cdot K_1(x,z)$ $\quad x = \begin{bmatrix} a \\ b \end{bmatrix}$ $\quad \begin{bmatrix} -2aa & -2ab \\ -2ab & -2bb \end{bmatrix}$

from Example before, we now have $-a < 0$,

So kernel is not valid

e) $K(x,z) = K_1(ax, bz) = e^{\cancel{x}\|ax - bz\|^2}$

The expression is always positive. Hence, Kernel is valid $\quad x = \begin{bmatrix} a \\ b \end{bmatrix}$ $\quad \begin{bmatrix} \cancel{baab} \; abba \\ abab \; abbb \end{bmatrix}$

f) $K(x,z) = \overset{>0}{K_1(x,z)} \cdot \overset{>0}{K_2(x,z)} > 0$

Kernel is valid $\quad \begin{bmatrix} a & b \\ a & b \end{bmatrix} \cdot \begin{bmatrix} b & a \\ b & a \end{bmatrix} = \begin{bmatrix} ba & aa \\ bb & ba \end{bmatrix}$

symmetry

g) $K(x,z) = f(x) f(z)$

That is a multiplication of functions resulted from $\langle \phi(x), \phi(z) \rangle = k(x,z)$

The Kernel is valid

h) $K(x,z) = K_3(\phi(x), \phi(z))$

Here the dimension of vectors does not influence the validity $\quad e^{1/2\|x-z\|^2} > 0$

i) $K(x,z) = p(K_1(x,z))$ Here the polinomial expression results in n number of multiplications and m numbers of summations of them. These multiplication have ~~are~~ always positive results.

Kernel is valid.

j) $K(x,z) = aK_1(x,z) - bK_2(x,z)$

We can assign b to be that large that the result of substraction ≤ 0. So, kernel is invalid.

k) $K(x,z) = -aK_1(x,z) - bK_2(x,z)$

In case we assign a and b to be positive numbers the expression becomes < 0. Kernel is not valid.

2. Task

a) $K(x,z) = x^T z + C$

$K(\varphi(x), \varphi(z)) = \varphi(x)^T \varphi(z) + C = x^T z + C + C =$

$= x^T z + 2C$

b) $K(x,z) = (x^T z + c)^d$, d - degree of polinomial

$K(\varphi(x), \varphi(z)) = (\varphi(x)^T \varphi(z) + c)^d =$

$= ((x^T z + c)^d + c)^d$

c) $K(x,z) = e^{-\varepsilon (x-z)^2}$

$K(\varphi(x), \varphi(z)) = e^{-\varepsilon(\varphi(x) - \varphi(z))^2} =$

$= e^{-\varepsilon(\varphi(x)^2 - 2\varphi(x)^T \varphi(z) + \varphi(z)^2)} =$

$= e^{-\varepsilon(e^{-\varepsilon(x-x)^2} - 2e^{-\varepsilon(x-z)^2} + e^{-\varepsilon(z-z)^2})} =$

$= e^{-\varepsilon(2 - 2e^{-\varepsilon(x-z)^2})} = e^{-2\varepsilon(1 - e^{-\varepsilon(x-z)^2})}$

3. Task

$$a_h(x) = g(x) = \frac{1}{1+e^{-w^Tx}}$$

Show that $a_h(x) = \tanh(x) = \dfrac{e^{w^Tx} - e^{-w^Tx}}{e^{w^Tx} + e^{-w^Tx}}$

$$\frac{1}{1+e^{-w^Tx}} \cdot \frac{e^{w^Tx}}{e^{w^Tx}} = \frac{e^{w^Tx}}{e^{w^Tx}+1} - \frac{1}{2} + \frac{1}{2} =$$

$$= \frac{2e^{w^Tx} - 1 - e^{w^Tx}}{2+2e^{w^Tx}} + \frac{1}{2} = \frac{1}{2}\left(\frac{2e^{w^Tx} - 1 - e^{w^Tx}}{1+e^{w^Tx}}\right) + \frac{1}{2} =$$

$$= \frac{1}{2}\left(\frac{e^{w^Tx} - 1}{1+e^{w^Tx}}\right) + 1/2 = \frac{1}{2}\left(\frac{e^{w^Tx}\left(e^{w^Tx/2} - e^{-w^Tx/2}\right)}{e^{w^Tx/2}\left(e^{w^Tx/2} + e^{-w^Tx/2}\right)}\right) + \frac{1}{2} =$$

$$= \frac{1}{2}\tanh\left(\frac{x}{2}\right) + \frac{1}{2}$$

7 task

Log reg. $h(x) = g(\theta^T x)$

$\theta_{ML} = \arg\max \prod_{i=1}^{m} p(y^{(i)}/x^{(i)}; \theta)$

If $p(\theta) \sim \mathcal{N}(0, \tau^2 I)$: $\theta_{map} = \arg\max p(\theta) \prod p(y^{(i)}/x^{(i)}; \theta)$

Proove: $\|\theta_{map}\|_2 \leq \|\theta_{ML}\|_2$

Proof from contradiction : $\|\theta_{map}\| > \|\theta_{ML}\|_2$.

Considering Normal distribution case we have its PDF : $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\boxed{\tau^2 = \sigma^2}$

$$p(\theta_{map}) = \frac{1}{\sqrt{2\pi\sigma^2 I}} e^{\frac{(\|\theta_{map}\|_2)^2}{2\sigma^2}}$$

$$p(\theta_{ml}) = \frac{1}{\sqrt{2\pi\sigma^2 I}} e^{\frac{-(\|\theta_{ml}\|_2)^2}{2\sigma^2}}$$

As we assume that $\|\theta_{map}\|_2 > \|\theta_{ml}\|_2$ in ~~th~~ conciderinq expessions above: $p(\theta_{ML}) > p(\theta_{map})$

So we could hence that:

$$p(\theta_{map}) \prod_{i=1}^{m} p(y^{(i)}/x^{(i)}, \theta_{map}) \ \cancel{\text{the}} <$$

$$< p(\theta_{ml}) \prod_{i=1}^{m} p(y^{(i)}/x^{(i)}, \theta_{map}), \text{ since } p(\theta_{ml}) > p(\theta_{map}).$$

~~θ_{ml} is originaly maximizes $\prod_{i=1}^{m} p(y^{(i)}/x^{(i)}; \theta)$~~

~~$p(\theta_{ml}) < p(\theta_{map})$ and Product of distributions~~

~~(considering PDF rule) will be higher in case tiny~~

This shows the contradiction to the primary condition of this task where $\theta_{map} = \arg\max p(\theta) \prod (y^{(i)}/x^{(i)}; \theta)$.

So, $\|\theta_{map}\|_2 \leq \|\theta_{ML}\|_2$