NEW YORK UNIVERSITY
TANDON SCHOOL OF ENGINEERING

Homework 3

Student name: *Zihan Zeng*

Course: *ECE 6143 Machine Learning* – Professor: *Yury Dvorkin*
Due date: *Oct 21st, 2021*

Problem 1

Solution.

(A) Mercer's Theorem Proof:

$$c^T K c = \sum_i \sum_j c_i c_j K_{ij} = \sum_i \sum_j c_i c_j \phi\left(x_i\right) \cdot \phi\left(x_j\right) = \left(\sum_i c_i \phi\left(x_i\right)\right) \cdot \left(\sum_i c_i \phi\left(x_i\right)\right)$$
$$= \left\|\sum_i c_i \phi\left(x_i\right)\right\|_2^2 \geq 0$$

(a) Proof:

$$\begin{aligned} k(x, \tilde{x}) &= \alpha k_1(x, \tilde{x}) + \beta k_2(x, \tilde{x}) \\ &= \left\langle \sqrt{\alpha}\phi_1(x), \sqrt{\alpha}\phi_1(\tilde{x}) \right\rangle + \left\langle \sqrt{\beta}\phi_2(x), \sqrt{\beta}\phi_2(\tilde{x}) \right\rangle \\ &= \left\langle \left[\sqrt{\alpha}\phi_1(x), \sqrt{\beta}\phi_2(x)\right], \left[\sqrt{\alpha}\phi_1(\tilde{x}), \sqrt{\beta}\phi_2(\tilde{x})\right] \right\rangle \end{aligned}$$

(b) Proof:

$$\begin{aligned} k(x, \tilde{x}) &= k_1(x, \tilde{x}) \times k_2(x, \tilde{x}) \\ &= (\phi_1(x) \cdot \phi_1(\tilde{x})) \times (\phi_2(x) \cdot \phi_2(\tilde{x})) \\ &= \left(\sum_{i=0}^{\infty} f_i(x) f_i(\tilde{x})\right) \times \left(\sum_{j=0}^{\infty} g_j(x) g_j(\tilde{x})\right) \\ &= \sum_{i,j} f_i(x) f_i(\tilde{x}) g_j(x) g_j(\tilde{x}) \\ &= \sum_{i,j} \left(f_i(x) g_j(x)\right) \left(f_i(\tilde{x}) g_j(\tilde{x})\right) \\ &= \langle \phi_3(x), \phi_3(\tilde{x}) \rangle \end{aligned}$$

(c) Proof:
We can know from the problem that f is any polynomial with positive coefficients, so the form of all items is similar to the following representation:

$$k(x,\tilde{x}) = \alpha k_1(x,\tilde{x}) + \beta k_2(x,\tilde{x}) + k_3(x,\tilde{x}) \times k_4(x,\tilde{x}) + \cdots$$

It can be seen that these terms are composed of the terms in (a) and (b), since the conditions in (a) and (b) are proved, so $k(x,\tilde{x}) = f(k_1(x,\tilde{x}))$ is also a Mercer kernel.

(d) Proof:
According to Taylor Expansion, we can get:

$$k(x,\tilde{x}) = e^{k_1(x,\tilde{x})} = \sum_{i=0}^{\infty} \frac{k_1(x,\tilde{x})^i}{i!}$$

Where $i >= 0$. Therefore, similar to (c), all items can be composed of items in (a) and (b). Therefore $k(x,\tilde{x}) = e^{k_1(x,\tilde{x})}$ is also a Mercer kernel.

(B) Proof:

Since the problem does not specify the norm, according to the convention, it is assumed to be the 2-norm. So that we can get:

$$\begin{aligned} K(x,y) &= e^{-\frac{1}{2}\|x-y\|^2} \\ &= e^{-\frac{1}{2}(x-y)^2} = e^{-\frac{1}{2}(x^2-2xy+y^2)} \quad = e^{-\frac{1}{2}x^2} \cdot e^{-\frac{1}{2}y^2} \cdot e^{xy} \end{aligned}$$

By Taylor expansion, we can get:

$$\begin{aligned} e^{xy} &= \sum_{n=0}^{\infty} \frac{(xy)^n}{n!} = 1 + xy + \frac{x^2y^2}{2!} + \cdots + \frac{x^ny^n}{n!} \\ &= \begin{bmatrix} 1 & \frac{1}{1!}x & \sqrt{\frac{1}{2!}}x^2 & \cdots \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1}{1!}y \\ \sqrt{\frac{1}{2!}}y^2 \\ \vdots \end{bmatrix} \end{aligned}$$

Then put this equation back to $K(x,y)$ and replace e^{xy}:

$$K(x,y) = e^{-\frac{1}{2}x^2} \cdot e^{-\frac{1}{2}y^2} \cdot e^{xy} = e^{-\frac{1}{2}x^2} \cdot \begin{bmatrix} 1 & \frac{1}{1!}x & \sqrt{\frac{1}{2!}}x^2 & \cdots \end{bmatrix} \cdot e^{-\frac{1}{2}y^2} \cdot \begin{bmatrix} 1 \\ \frac{1}{1!}y \\ \sqrt{\frac{1}{2!}}y^2 \\ \vdots \end{bmatrix}$$

So an explicit formula for φ is:

$$\varphi(x) = e^{-\frac{1}{2}x^2} \cdot \begin{bmatrix} 1 \\ \frac{1}{1!}x \\ \sqrt{\frac{1}{2!}}x^2 \\ \vdots \end{bmatrix}^T, \varphi(y) = e^{-\frac{1}{2}y^2} \cdot \begin{bmatrix} 1 \\ \frac{1}{1!}y \\ \sqrt{\frac{1}{2!}}y^2 \\ \vdots \end{bmatrix}$$

Problem 2

Solution. According to the code in svm.m and svkernel.m, modify different kernels respectively, and then modify the C value and the value of sigma or polynomial order.

(*a*) Linear Kernel
The value that needs to be modified here is C.

C	-5	-4	-3	-2	-1	0	1	2	3	4	5	inf
Error	30	27	23	27	25	27	26	25	22	26	23	23

(*b*) Polynomial Kernel
The values that need to be modified here are polynomial order and C.

C \Polynomial Order	0	1	2	3	4	5	6	7	8	9	10
-5	25	25	24	24	28	27	22	27	22	25	25
-4	28	27	21	25	23	26	28	28	27	23	26
-3	23	25	27	34	24	28	26	22	26	22	22
-2	24	25	22	22	26	25	27	23	20	26	25
-1	24	23	25	27	21	22	27	26	24	24	24
0	22	26	19	28	24	25	29	28	26	24	23
1	28	29	23	22	30	27	26	27	28	21	28
2	24	21	20	26	26	26	23	27	28	26	27
3	22	25	27	30	21	25	23	25	25	25	23
4	23	21	23	28	22	27	23	26	27	25	25
5	22	25	24	21	27	25	19	26	27	24	26
inf	23	28	28	25	23	20	24	22	24	27	27

(*c*) RBF Kernel
The values that need to be modified here are sigma and C.

C \Sigma	-5	-4	-3	-2	-1	0	1	2	3	4	5
-5	28	26	25	25	29	28	27	29	22	21	23
-4	24	23	30	23	24	23	29	28	24	23	27
-3	28	19	20	23	28	22	28	25	26	26	25
-2	29	23	19	23	26	27	21	28	21	25	28
-1	22	22	22	27	29	23	26	20	19	26	28
0	22	24	19	21	25	25	22	20	20	20	24
1	26	22	20	25	27	23	28	27	20	22	29
2	21	22	27	25	22	26	25	22	25	21	27
3	24	23	27	22	23	27	25	27	28	20	25
4	26	29	19	25	22	27	20	22	21	21	23
5	22	23	26	27	23	20	19	26	27	33	20
inf	25	22	21	25	22	25	24	27	26	22	26

Problem 3

Solution. According to the question, we can know that the random variables satisfie the situation of independent and identical distribution, so we can get the likelihood function:

$$L(\alpha) = \prod_{i=1}^{n} f\left(x_i; \alpha\right)$$

Then

$$L(\alpha) = \prod_{i=1}^{n} f\left(x_i; \alpha\right) = \prod_{i=1}^{n} \alpha^{-2} x_i e^{-\frac{x_i}{\alpha}} = \alpha^{-2n} \left(\prod_{i=1}^{n} x_i\right) e^{\left(\frac{-\sum_{i=1}^{n} x_i}{\alpha}\right)}$$

Because the computer has precision problems when dealing with floating-point numbers, values that are too small cannot be represented. Therefore the log-likelihood function is used here:

$$\ell(\alpha) = \ln L(\alpha) = -2n \ln \alpha + \sum_{i=1}^{n} \ln x_i - \frac{\sum_{i=1}^{n} x_i}{\alpha}$$

Then taking derivative of α:

$$\frac{d\ell(\alpha)}{d\alpha} = \frac{-2n}{\alpha} + \frac{\sum_{i=i}^{n} x_i}{\alpha^2}$$

And we let this formula equal to zero, and then we can solve for α:

$$\alpha = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{\bar{x}}{2}$$

According to the problem, we know that $x_1 = 0.25, x_2 = 0.75, x_3 = 1.50, x_4 = 2.5, x_5 = 2.0$, and we put these values into the formula:

$$\alpha = \frac{0.25 + 0.75 + 1.50 + 2.50 + 2.0}{2 \cdot 5} = 0.70$$