Problem 1

(A)

c is a $n\times 1$ matrix

c^T is a $1\times n$ matrix

K is a $n\times n$ matrix and $K_{i,j}=k(x_i,x_j)$

$$\mathrm{c}^{\mathrm{T}}Kc=\begin{bmatrix}c_1 & c_2 & \cdots & c_n\end{bmatrix}\begin{bmatrix}k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n)\\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n)\\ \vdots & \vdots & \ddots & \vdots\\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n)\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix}$$

$$=\begin{bmatrix}\sum_{i=1}^{n}c_iK_{i,1} & \sum_{i=1}^{n}c_iK_{i,2} & \cdots & \sum_{i=1}^{n}c_iK_{i,n}\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix}$$

$$=c_1\sum_{i=1}^{n}c_iK_{i,1}+c_2\sum_{i=1}^{n}c_iK_{i,2}+\cdots+c_n\sum_{i=1}^{n}c_iK_{i,n}$$

$$=\sum_{j=1}^{n}c_j\sum_{i=1}^{n}c_iK_{i,j}$$

$$\begin{aligned}=\quad & c_1c_1K_{1,1}+c_1c_2K_{1,2}+\cdots+c_1c_nK_{1,n}\\ &+c_2c_1K_{2,1}+c_2c_2K_{2,2}+\cdots+c_2c_nK_{2,n}\\ &\quad\vdots\qquad\quad\vdots\qquad\quad\vdots\qquad\vdots\\ &+c_nc_1K_{n,1}+c_nc_2K_{n,2}+\cdots+c_nc_nK_{n,n}\end{aligned}$$

$$\begin{aligned}\because\ & c_1c_2K_{1,2}+c_2c_1K_{2,1}=2c_1c_2K_{1,1}K_{2,2}\\ & c_1c_3K_{1,3}+c_3c_1K_{3,1}=2c_1c_3K_{1,1}K_{3,3}\\ &\quad\vdots\qquad\qquad\vdots\qquad\qquad\vdots\end{aligned}$$

The elements on both sides symmetrical to the diagonal can be calculated in this way.

And $K_{i,i}=\phi^2(x_i)$.

$$\therefore\ \mathrm{c}^{\mathrm{T}}Kc=(\sum_{i=1}^{n}c_i\phi(x_i))^2\geq 0$$

So the kernel matrix K is positive semi-definite.

$a)$

For $k_1(x,\tilde{x})$, $\mathrm{c}^{\mathrm{T}}Kc=(\sum_{i=1}^{n}c_i\phi(x_i))^2\geq 0$

For $k_2(x,\tilde{x})$, $\mathrm{c}^{\mathrm{T}}Kc=(\sum_{i=1}^{n}c_i\phi(x_i))^2\geq 0$

For $k(x,\tilde{x})$, $\alpha,\beta\geq 0$, $\mathrm{c}^{\mathrm{T}}Kc=\alpha(\sum_{i=1}^{n}c_i\phi(x_i))^2+\beta(\sum_{i=1}^{n}c_i\phi(x_i))^2\geq 0$

$b)$

For $k_1(x,\tilde{x})$, $c^T Kc = (\sum_{i=1}^{n} c_i \phi(x_i))^2 \geq 0$

For $k_2(x,\tilde{x})$, $c^T Kc = (\sum_{i=1}^{n} c_i \phi(x_i))^2 \geq 0$

For $k(x,\tilde{x})$, $c^T Kc = (\sum_{i=1}^{n} c_i \phi(x_i))^2 (\sum_{i=1}^{n} c_i \phi(x_i))^2 \geq 0$

$c)$

For $k_1(x,\tilde{x})$, $c^T Kc = (\sum_{i=1}^{n} c_i \phi(x_i))^2 \geq 0$

Set a_i as the index of each term in any polynomial

$$\sum_{i=1}^{n} (c_i \phi^{a_i}(x_i))^2 \geq 0$$

set A_i as the as the positive coefficients of each term in any polynomial

For $k(x,\tilde{x})$, $c^T Kc = \sum_{i=1}^{n} A_i (c_i \phi^{a_i}(x_i))^2 \geq 0$

$d)$

For $k(x,\tilde{x})$, $c^T Kc =$

$$\begin{aligned}
& c_1c_1K_{1,1} + c_1c_2K_{1,2} + \cdots + c_1c_nK_{1,n} \\
& + c_2c_1K_{2,1} + c_2c_2K_{2,2} + \cdots + c_2c_nK_{2,n} \\
& \quad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
& + c_nc_1K_{n,1} + c_nc_2K_{n,2} + \cdots + c_nc_nK_{n,n} \\
= \; & c_1c_1e^{k_1(x_1,x_1)} + c_1c_2e^{k_1(x_1,x_2)} + \cdots + c_1c_ne^{k_1(x_1,x_n)} \\
& + c_2c_1e^{k_1(x_2,x_1)} + c_2c_2e^{k_1(x_2,x_2)} + \cdots + c_2c_ne^{k_1(x_2,x_n)} \\
& \quad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
& + c_nc_1e^{k_1(x_n,x_1)} + c_nc_2e^{k_1(x_n,x_2)} + \cdots + c_nc_ne^{k_1(x_n,x_n)} \\
= \; & c_1c_1(\sum_{j=0}^{\infty} \frac{k_1^j(x_1,x_1)}{j!}) + c_1c_2(\sum_{j=0}^{\infty} \frac{k_1^j(x_1,x_2)}{j!}) + \cdots + c_1c_n(\sum_{j=0}^{\infty} \frac{k_1^j(x_1,x_n)}{j!}) \\
& + c_2c_1(\sum_{j=0}^{\infty} \frac{k_1^j(x_2,x_1)}{j!}) + c_2c_2(\sum_{j=0}^{\infty} \frac{k_1^j(x_2,x_2)}{j!}) + \cdots + c_2c_n(\sum_{j=0}^{\infty} \frac{k_1^j(x_2,x_n)}{j!}) \\
& \quad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
& + c_nc_1(\sum_{j=0}^{\infty} \frac{k_1^j(x_n,x_1)}{j!}) + c_nc_2(\sum_{j=0}^{\infty} \frac{k_1^j(x_n,x_2)}{j!}) + \cdots + c_nc_n(\sum_{j=0}^{\infty} \frac{k_1^j(x_n,x_n)}{j!}) \\
= \; & \sum_{j=0}^{\infty} \frac{(\sum_{i=1}^{n} c_i \phi_1^j(x_i))^2}{j!} \geq 0
\end{aligned}$$

(B)

Assume x and y are 2 dimensional vectors

$$
\begin{aligned}
K(x,y) &= e^{-\frac{1}{2}\|x-y\|^2} \\
&= e^{-\frac{1}{2}((x_1-y_1)^2+(x_2-y_2)^2)} \\
&= e^{-\frac{1}{2}(x_1^2+x_2^2+y_1^2+y_2^2-2x_1y_1-2x_2y_2)} \\
&= e^{-\frac{1}{2}\|x\|^2} e^{-\frac{1}{2}\|y\|^2} e^{x^T y} \\
&= e^{-\frac{1}{2}\|x\|^2} e^{-\frac{1}{2}\|y\|^2} \sum_{j=0}^{\infty} \frac{(x^T y)^j}{j!}
\end{aligned}
$$

$$So\ \phi(x) = e^{-\frac{1}{2}\|x\|^2} \sum_{j=0}^{\infty} \frac{x^j}{\sqrt{j!}}$$

Problem 2

①Firstly, 70% of the data in the dataset are randomly divided for training and 30% for testing.

Here is the random classification matlab code:

```
n=size(X,1);
Y(Y==0)=-1;
n_train = 65;
randnum=randperm(size(X,1));
x_train = X(randnum(1:n_train),:);
y_train = Y(randnum(1:n_train),:);
size(x_train)
size(y_train)
x_test = X(randnum(n_train+1:end),:);
y_test = Y(randnum(n_train+1:end),:);
```

②Secondly, write the code to show the errors in test set with different C and p (for Polynomial Kernel) or σ （for RBF Kernel）.

Here is the code:

```
global p
ker = 'rbf';
Cs = [0,1,2,3,4,inf];
kers = ['linear','poly','rbf'];
C=inf;
errors=zeros(6,3,9);
for i=1:6
    for j = 1:3
        for k = 1:9
            p=k;
```

```
            C=Cs(i);
            ker = kers(1,j);
            C;
            ker;
            [nsv, alpha, b0] = svc(x_train, y_train, ker, C);
            err = svcerror(x_train,y_train,x_test,y_test,ker,alpha,b0);
            errors(i,j,k)=err;
            err;
        end
    end
end
errors
```

③We can list the relationship between errors and C and p/σ.

For Linear Kernel,

C	0	1	2	3	4	inf
Errors	30	2	3	1	1	5

We can see that when C=3 or 4, there are least errors.

For Polynomial Kernel,

C	0	1	2	3	4	inf
p=1, errors	30	1	1	1	1	15
p=2, errors	30	2	1	2	2	5
p=3, errors	30	3	3	15	3	3
p=4, errors	30	1	1	1	1	15
p=5, errors	30	5	5	5	5	5
p=6, errors	30	15	15	15	15	15
p=7, errors	30	5	5	5	5	5
p=8, errors	30	15	15	15	15	15
p=9, errors	30	3	3	15	3	3

We can see that when C=2 and p=2, there are least errors.

For RBF Kernel,

C	0	1	2	3	4	inf
σ=1, errors	30	2	3	2	2	5
σ=2, errors	30	4	2	2	1	5
σ=3, errors	30	4	4	2	2	3
σ=4, errors	30	5	4	4	4	3
σ=5, errors	30	5	5	4	4	3
σ=6, errors	30	5	5	5	5	3
σ=7, errors	30	5	4	4	4	3
σ=8, errors	30	5	4	5	4	3
σ=9, errors	30	5	5	5	5	3

We can see that when C=4 and σ=2, there are least errors.

Problem 3

The likelihood of $X = \{x_1, \cdots, x_N\}$ under IID assumptions is:

$$p(x \mid \alpha) = p(x_1, \cdots, x_N \mid \alpha) = \prod_{i=1}^{N} p_i(x_i \mid \alpha) = \prod_{i=1}^{N} p(x_i \mid \alpha)$$

Learn joint distribution $p(x_i \mid \alpha)$ by maximum likelihood:

$$\alpha^* = \arg\max_\alpha \prod_{i=1}^{N} f(x_i \mid \alpha) = \arg\max_\alpha \ln \prod_{i=1}^{N} f(x_i \mid \alpha)$$

From the information given, we can know that

$$f(x_1 \mid \alpha) = \alpha^{-2} x_1 e^{-\frac{x_1}{\alpha}} = \alpha^{-2} 0.25 e^{-\frac{0.25}{\alpha}}$$

$$f(x_2 \mid \alpha) = \alpha^{-2} x_2 e^{-\frac{x_2}{\alpha}} = \alpha^{-2} 0.75 e^{-\frac{0.75}{\alpha}}$$

$$f(x_3 \mid \alpha) = \alpha^{-2} x_3 e^{-\frac{x_3}{\alpha}} = \alpha^{-2} 1.50 e^{-\frac{1.50}{\alpha}}$$

$$f(x_4 \mid \alpha) = \alpha^{-2} x_4 e^{-\frac{x_4}{\alpha}} = \alpha^{-2} 2.50 e^{-\frac{2.50}{\alpha}}$$

$$f(x_5 \mid \alpha) = \alpha^{-2} x_5 e^{-\frac{x_5}{\alpha}} = \alpha^{-2} 2.00 e^{-\frac{2.00}{\alpha}}$$

$$\prod_{i=1}^{N} f(x_i \mid \alpha) = \prod_{i=1}^{5} f(x_i \mid \alpha) = \alpha^{-10} x_1 x_2 x_3 x_4 x_5 e^{-\frac{\sum_{i=1}^{5} x_i}{\alpha}}$$

$$\alpha^* = \arg\max_\alpha \prod_{i=1}^{5} f(x_i \mid \alpha) = \arg\max_\alpha \ln \prod_{i=1}^{5} f(x_i \mid \alpha) = \arg\max_\alpha (\ln(x_1 x_2 x_3 x_4 x_5) - \frac{\sum_{i=1}^{5} x_i}{\alpha} - 10 \ln \alpha)$$

$$\frac{\partial \prod_{i=1}^{5} f(x_i \mid \alpha)}{\partial \alpha} = \frac{\sum_{i=1}^{5} x_i}{\alpha^2} - \frac{10}{\alpha} = 0$$

$$\frac{7}{\alpha^2} - \frac{10}{\alpha} = 0$$

$$\alpha = 0.7$$