Assignment 1 : CS-E4830 Kernel Methods in Machine

Yi Zhang 602466

**Question 1. If K is a positive semidefinite kernel matrix, then all its entries must be positive.**

**Answer:**

False,  For example $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.,not all entries are positive.

**Question 2.**

**Answer:**

True, for any none-zero vector v:

$$v^T K v = v^T (aK_1 + bK_2) v$$

$$v^T K v = a v^T K_1 v + b v^T K_2 v$$

as we know,

$$a v^T K_1 v >= 0, b v^T K_2 v >= 0$$

so

$$v^T K v >= 0$$

**Question 3**

**Answer:**

False, for example $\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} - \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ K is not a valid kernel matrix.

**Parzen window classifier**

**Question 4**

$$h(x) = sign(\left\| \phi(x) - \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i) \right\|^2 - \left\| \phi(x) - \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i) \right\|^2 )$$

$$h(x) = sign(k(x,x) - \frac{2}{m_-} \sum_{i \in I^-} k(x, x_i) + \frac{1}{m^2_-} \sum_{i,j \in I^-} k(x_i, x_j) - k(x,x) + \frac{2}{m_+} \sum_{i \in I^+} k(x, x_i) - \frac{1}{m^2_+} \sum_{i,j \in I^+} k(x_i, x_j))$$

$$h(x) = sign(\frac{1}{m^2_-} \sum_{i,j \in I^-} k(x_i, x_j) - \frac{1}{m^2_+} \sum_{i,j \in I^+} k(x_i, x_j) - \frac{2}{m_-} \sum_{i \in I^-} k(x, x_i) + \frac{2}{m_+} \sum_{i \in I^+} k(x, x_i))$$

$$h(x) = sign(\frac{1}{m_+} \sum_{i \in I^+} k(x, x_i) - \frac{1}{m_-} \sum_{i \in I^-} k(x, x_i) + b)$$

and $b = \frac{1}{2m_-^2} \sum_{i,j \in I^-} k(x_i, x_j) - \frac{1}{2m_+^2} \sum_{i,j \in I^+} k(x_i, x_j)$

We define a vector $a_i$

$$a_i \begin{cases} \frac{1}{m_+} & \text{if y=+1} \\ \frac{1}{m_-} & \text{if y=-1} \end{cases}$$

Then it could be written as

$$h(x) = sign(\sum a_i k(x, x_i) + b)$$

**Question 5:**

**answer:**

**see the code "gauss_kernel.m"**

```
function [kernel] = gauss_kernel(X, Z, sigma)
%X = X_train;
%Z = X_test;
%diag(X*X')return 800*1   *1*200           800*1   *           1*200
% (X-Z)^2
n1= size(X, 1);
n2 = size(Z, 1);
distance = diag(X*X')*ones(1, n2) + ones(n1, 1)*diag(Z*Z')'-2*X*Z';

%gauss_kernel
kernel = exp(-distance./(2 * sigma ^2));

end
```

**Question 6:**

**see the code "h_func.m"**

```
function [y_predict] = h_func(K_train, K_traintest, y_train)
%find the y_train label
pos_ind = find(y_train == 1);
neg_ind = find(y_train == -1);
num_pos = length(pos_ind);
num_neg = length(neg_ind);

%size of training and test examples
[num_train,num_test] = size(K_traintest);

% train
ai = zeros(num_train,1);
ai(pos_ind)=1/num_pos;
ai(neg_ind)=-1/num_neg;

b = 1/(2*num_neg^2)*sum(sum(K_train(neg_ind,neg_ind)))-1/(2*num_pos^2) * sum(sum(K_train(pos_ind,pos_ind)));

%predict
y_predict = sign(ai'* K_traintest + repmat(b,1,num_test));
end
```

**Question 7**

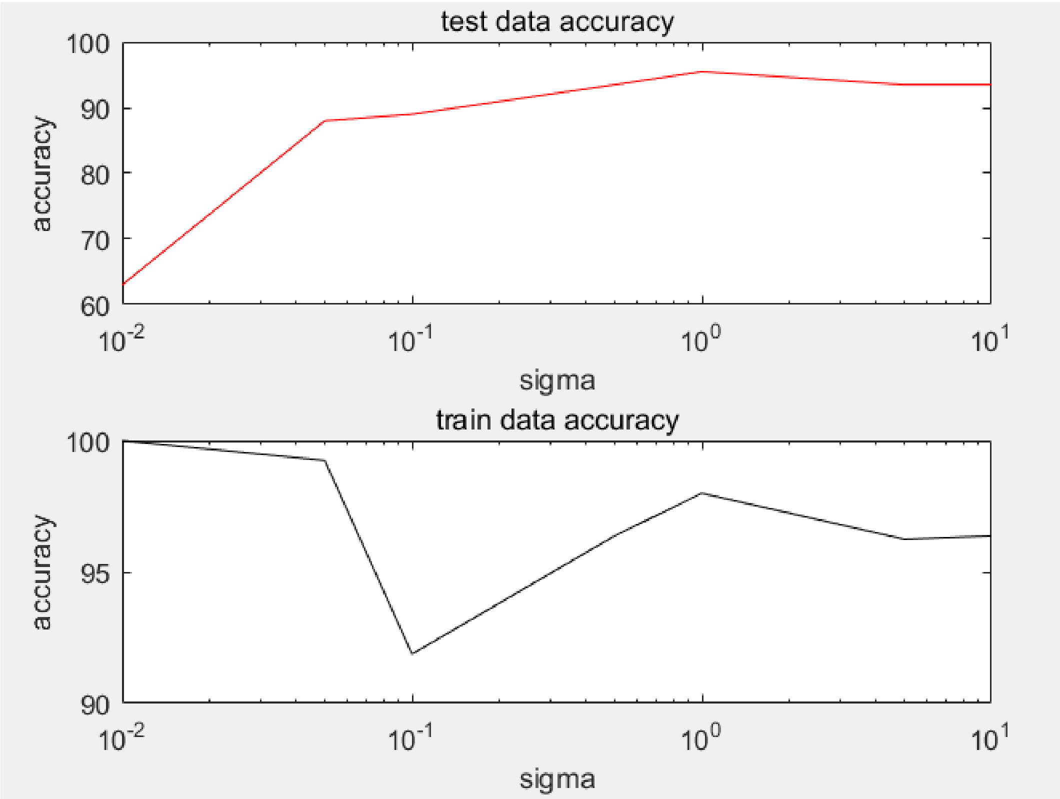


**Figure 1**

As we could see from the Figure1, in small sigma region, test accuracy is low and train data accuracy is high because of the overfitting, however, in big sigma region, the test accuracy is relatively large.