Question 1

a. $c_r^i = [-\frac{r}{2}, \frac{r}{2}]$ is the interval of ith dimension of cube C_r^D.
Because $Vol_D A = \int_A dx_1...dx_D$,

$$\begin{aligned} Vol_D(C_r^D) &= \int_{C_r^D} dx_1...dx_D \\ &= \int_{c_r^D} ... \int_{c_r^1} dx_1...dx_D \\ &= \prod_{i=1}^{D} \int_{c_r^i} dx_i \\ &= \prod_{i=1}^{D} r = r^D \end{aligned} \tag{1}$$

b. Since $A_{\epsilon,r}^D = \{xC_r^D | x \notin C_\epsilon^D\} = C_r^D - C_{r-\epsilon}^D$,

$$\begin{aligned} Vol_D(A_{\epsilon,r}^D) &= \int_{A_{\epsilon,r}^D} dx_1...dx_D \\ &= \int_{C_r^D - C_{r-\epsilon}^D} dx_1...dx_D \\ &= \int_{C_r^D} dx_1...dx_D - \int_{C_{r-\epsilon}^D} dx_1...dx_D \\ &= r^D - (r-\epsilon)^D \end{aligned} \tag{2}$$

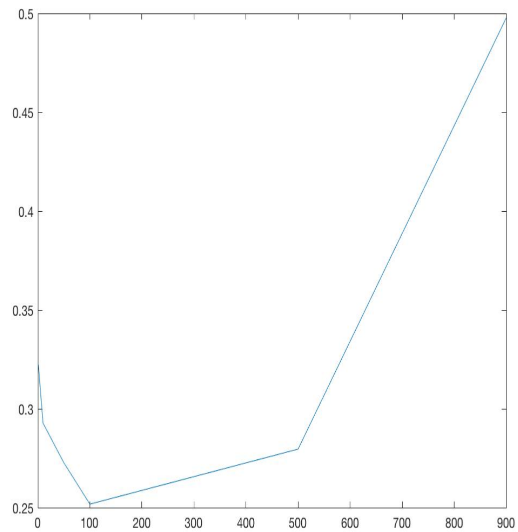
Then $\frac{Vol_D(A_{\epsilon,r}^D)}{Vol_D(C_r^D)} = \frac{r^D - (r-\epsilon)^D}{r^D} = 1 - (1 - \frac{\epsilon}{r})^D$

c. When $D = 10$, if $\epsilon = \frac{r}{10}$, $\frac{Vol_D(A_{\epsilon,r}^D)}{Vol_D(C_r^D)} = 65.13\%$.
Because $0 < \epsilon < r$, $1 - \frac{\epsilon}{r} < 1$ and $\frac{r}{r-\epsilon} > 1$

$$\begin{aligned} \frac{\partial}{\partial D}(\frac{Vol_D(A_{\epsilon,r}^D)}{Vol_D(C_r^D)}) &= -\frac{\partial}{\partial D}(1 - \frac{\epsilon}{r})^D \\ &= ln(\frac{r}{r-\epsilon})(1 - \frac{\epsilon}{r}) > 0 \end{aligned} \tag{3}$$

It means that the ratio $\frac{Vol_D(A_{\epsilon,r}^D)}{Vol_D(C_r^D)}$ monotonically increases as D get larger.

Question 2



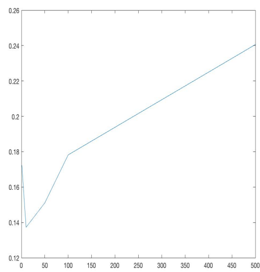
The kNN of best performance among this kNN set,$\{1, 10, 50, 100, 500, 900\}$, is at kNN = 100. When kNN is too small, the model will be underfitting; when kNN is too large, the model will be overfitting.

The code will be attached at the end.

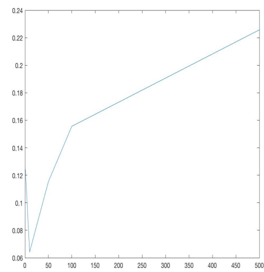
Question 3

Basing on the results from 10 epoches (below only shows 4 of 10 epoches), when the kNN is smaller 100, the loss will be small. Sometimes when kNN is 1, loss will be relatively large, because a large amount testing points is at boundary of the partitions.

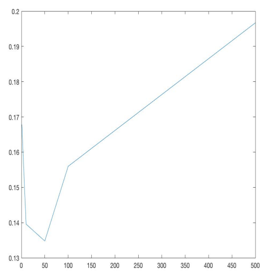
The code will be attached at the end.



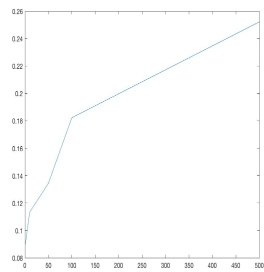
(a) Epoch 1



(b) Epoch 2



(c) Epoch 3



(d) Epoch 4

Question 4

a. $E = \{x \in \mathbb{R}^d | w^T x = 0\}$, it means E is nullspace of w^T, $E = Null(w^T)$
Because $w^T \in \mathbb{R}^d$ and $w^T \neq 0$

$$dim(E) = dim(\mathbb{R}^d) - dim(w^T) = d - 1 \tag{4}$$

b. Part (a) has proved that $S = \{x \in \mathbb{R} | w^T x = 0\}$ has rank $d - 1$. We can rewrite S as below

$$\begin{aligned} S &= \{(x+v) \in \mathbb{R}^d | w^T(x+v) = 0, \text{for some } v \in \mathbb{R}^d \text{ s.t. } w^T v = -b\} \\ &= \{(x+v) \in \mathbb{R}^d | w^T x = b, \text{for some } v \in \mathbb{R}^d \text{ s.t. } w^T v = -b\} \end{aligned} \tag{5}$$

Now $E = \{x \in \mathbb{R}^d | w^T x = b\}$ is isomorphic to S by **a translation**, that is

$$E = \{x \in \mathbb{R}^d | x = y - v \text{ for } y \in S \text{ and some } v \in \mathbb{R}^d \text{ s.t. } w^T v = -b\}$$

Thus E has same dimension as S, that is $d - 1$.

0.1 Code of Question 2

```
  load kNN_ClassifierSyntheticData
% Labels: 1x1000 lebels
% X : 1000X2 Data point


%% Random sample data

XLabels = [X Labels.'];

testidx = randperm(1000,100);

test100 = XLabels(testidx,:) ;

trainidx = setdiff([1:1000],testidx);

%training set
train900 = XLabels(trainidx,:) ;

samX= train900(:,1:2);
samL= train900(:,3);

%initiate data
kNNset = [1 10 50 100 500 900];
Loss100 =[];

%% Model builder

for kNN = kNNset
   Mdl = fitcknn(samX,samL,'NumNeighbors',kNN);
   Loss100 = [Loss100 loss(Mdl,test100(:,1:2),test100(:,3))];
end
```

```
load SalinasA_gt
load SalinasA


%% Random sample data

lineSal = reshape(salinasA,[],224);
lineGT = reshape(salinasA_gt,[],1);

XLabels = [lineSal lineGT];

%test set
testidx = randperm(length(XLabels),100);

test100 = XLabels(testidx,:) ;
```

```
%training set
trainidx = setdiff([1:length(XLabels)],testidx);

trainSet = XLabels(trainidx,:) ;

samX= trainSet(:,1:224);
samL= trainSet(:,225);

%initiate data
kNNset = [1 10 50 100 500];
Loss100 =[];

%% Model builder

for kNN = kNNset
    Mdl = fitcknn(samX,samL,'NumNeighbors',kNN);
    Loss100 = [Loss100 loss(Mdl,test100(:,1:224),test100(:,225))];
end
```