

# IE 6318 Data Mining and Analytics

## Homework 2

### Question 1:

```
data = importdata('iris.txt');
training = data([1:40,51:90,101:140],1:4);
trainingL = data([1:40,51:90,101:140],5);
testing = data([41:50, 91:100, 141:150], 1:4);
testingL = data([41:50, 91:100, 141:150], 5);
feat = data(:,1:13); % feature matrix
label = data(:,14); % class label vector
%size(training)
%size(testing)
```

### Question 2:

```
for k = [3 5 7]
    for r = [1 2 5]
        for i = 1:30
            A = training;
            B = testing(i,:);
            dist = minkowski(A, B, r);
            [sorted, index] = sort(dist);
            knnindex = index(:,1:k).';
            knnclass = trainingL(knnindex);
            pred(i, :) = mode(knnclass);

        end

        confu = confusionmat(testingL, pred)
        x = 0;
        for i = 1:30
            if pred(i) == testingL(i)
                x = x+1;
            end
        end
        acc = x/30*100
        x = 0;
        for i = 1:10
            if pred(i) == testingL(i)
                x = x+1;
            end
        end
        acc1 = x/10*100
        x = 0;
        for i = 11:20
            if pred(i) == testingL(i)
                x = x+1;
            end
        end
        acc2 = x/10*100
    end
end
```

```

        x = 0;
        for i = 21:30
            if pred(i) == testingL(i)
                x = x+1;
            end
        end
        acc3 = x/10*100
    end
end

```

confu =

10	0	0
0	10	0
0	0	10

acc =

100

acc1 =

100

acc2 =

100

acc3 =

100

confu =

10	0	0
0	10	0
0	0	10

acc =

100

```
acc1 =  
    100  
  
acc2 =  
    100  
  
acc3 =  
    100  
  
confu =  
    10    0    0  
    0    10   0  
    0    0    10  
  
acc =  
    100  
  
acc1 =  
    100  
  
acc2 =  
    100  
  
acc3 =  
    100  
  
confu =  
    10    0    0  
    0    10   0  
    0    0    10
```

```
acc =
```

```
    100
```

```
acc1 =
```

```
    100
```

```
acc2 =
```

```
    100
```

```
acc3 =
```

```
    100
```

```
confu =
```

```
    10    0    0
     0   10    0
     0    0   10
```

```
acc =
```

```
    100
```

```
acc1 =
```

```
    100
```

```
acc2 =
```

```
    100
```

```
acc3 =
```

```
    100
```

confu =

10	0	0
0	10	0
0	0	10

acc =

100

acc1 =

100

acc2 =

100

acc3 =

100

confu =

10	0	0
0	10	0
0	0	10

acc =

100

acc1 =

100

acc2 =

100

acc3 =

100

confu =

10	0	0
0	10	0
0	0	10

acc =

100

acc1 =

100

acc2 =

100

acc3 =

100

confu =

10	0	0
0	10	0
0	0	10

acc =

100

```

acc1 =

    100

acc2 =

    100

acc3 =

    100

```

```

for K = [3 5 7]
for Dorder = [1 2 5]
    for ifold = 1:N
        %----prepare cross-validation training and testing dataset---%
        idx_test = ifold; % index for testing fold
        idx_train = setdiff(1:N, ifold); % index for training folds
        Dtest = []; Ltest = []; % initialize testing data and label
        Dtrain = []; Ltrain = []; % initialize testing data and label

        %---construct the training and testing dataset for the ith fold cross
        validatoin
        for iC = 1:length(C)
            cl = C(iC);
            dtest = eval(['data_nfold.class',num2str(iC), '.fold',
num2str(ifold)]);
            Dtest = [Dtest; dtest];
            Ltest = [Ltest; cl*ones(size(dtest,1), 1)];

            for itr = 1:length(idx_train)
                idx = idx_train(itr);
                dtrain = eval(['data_nfold.class',num2str(iC), '.fold',
num2str(idx)]);
                Dtrain = [Dtrain; dtrain];
                Ltrain = [Ltrain; cl*ones(size(dtrain,1), 1)];
            end
        end
        Lpred = Ltest; % predicted label for testing dataset

        %---Calculate Classification Accuracy-----%
        acc = sum(Lpred==Ltest)/length(Ltest);
    end
end
end

```

```

function H = myknn(A,B,C,r,k)
%A training data set
%B test data set
%C label of training data set
p =size(A,1);
q =size(B,1);
for z= 1:q
    D =abs(minko_dist123(A,B(z,:),r));
    [F,I]= sort(D);
    G = I(2:k+1,1);
    V = C(G,1);
    m=sum(V==1);
    s= sum(V==2);
    l= sum(V==3);
    if (m>=s) && (m>=l)
        H(z,1) =1;
    elseif (s>=m) && (s>=l)
        H(z,1) =2;
    else(l>=m) && (l>=s);
        H(z,1) =3;
    end
end
end

```

KNN=3	CLASSIFICATION ACCURACY
1	100%
2	100%
5	100%
KNN=5	CLASSIFICATION ACCURACY
1	100%
2	100%
5	100%
KNN=7	CLASSIFICATION ACCURACY
1	100%
2	100%
5	100%



2  
2  
3  
2  
2  
2  
2  
2  
2  
2  
3  
3  
3  
3  
3  
3  
2  
3  
3  
3  
2  
2  
2  
2  
2  
2  
2  
2

2

2

3

3

3

3

3

3

3

3

3

3

3

2

3

2

2

2

2

3

2

2

3

3

3

3

3

3

3  
3  
3  
3  
2  
2  
2  
3  
2  
2  
2  
2  
2  
2  
2  
3  
3  
3  
3  
3  
3  
3  
3  
3  
2  
2  
2  
2

2  
2  
2  
2  
2  
2  
3  
3  
3  
3  
3  
3  
3  
3  
3  
3

### Question 3:

```
function z = decissionontree(C,D)
p =size(C,1);
for i=1:p
    if D(i,1) >= 1.7 && C(i,1)>4.3
        z(i,1) =3;
    else
        z(i,1) =2;
    end
end
end
```

Classification Accuracy = 98%  
Specificity =  $TN/(TN+FP) = 49/(49+5) = 90.74\%$   
Sensitivity =  $TP/(TP+FN) = 45/(46) = 97.8\%$

Confusion Matrix  
45 5

