In [115]:
```
import init
import mnistdata as mylib
import numpy as np
from matplotlib import pyplot
import matplotlib as mpl
import matplotlib.pyplot as plt
import math as math
import preprocessing as pp
import knnfeatures as kf
import datetime
import sys


def show(image,image2='None'):
    fig = pyplot.figure()
    ax = fig.add_subplot(1,1,1)
    imgplot = ax.imshow(image,cmap=mpl.cm.Greys)
    imgplot.set_interpolation('nearest')
    ax.xaxis.set_ticks_position('top')
    ax.yaxis.set_ticks_position('left')
    if(image2!='None'):
```

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

```
    pyplot.show()

def euclideanDistance(x1, x2):
    dis = sum(pow(x1-x2,2))
    return np.sqrt(dis)

def oneNormDistance(x1, x2):
    dis = sum(abs(x1-x2))
    return dis

def infNormDistance(x1,x2):
    dis = np.zeros(x1.size)
    for i in range(x1.size):
        dis[i] = max(x1[i],x2[i])
    return(sum(dis))

def mahalonobisDist(x1,x2,invS):
    diff = np.matrix(x1-x2)
    print(diff)
    dis = (diff*invS)*diff.T
    return dis

def getFeatureVector(Image):
    #pre-processed Image
```

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

```
    featureVector.append(kf.averagemass(ppImage))
    #feature 2,3: Number of Left/Right pixels
    lrmass = kf.averagelrmass(ppImage)
    featureVector.append(lrmass[0])
    featureVector.append(lrmass[1])
    #feature 4,5: Number of Top/Bottom pixels
    tbmass = kf.averagetbmass(ppImage)
    featureVector.append(tbmass[0])
    featureVector.append(tbmass[1])
    #feature 6: avgHorizontalStroke
    featureVector.append(kf.avgHorizontalStroke(ppImage))
    #feature 7: avgVerticalStroke
    featureVector.append(kf.avgVerticalStroke(ppImage))
    featureVector.append(kf.transitions(ppImage))
    featureVector.append(kf.topBoundaryTouch(ppImage))
    featureVector.append(kf.bottomBoundaryTouch(ppImage))
    featureVector.append(kf.leftBoundaryTouch(ppImage))
    featureVector.append(kf.rightBoundaryTouch(ppImage))
    featureVector.append(kf.aspectRatio(ppImage))
    featureVector.append(kf.avgDistanceFromImageCenter(ppImage))
    featureVector.append(kf.ySymmetric(ppImage))
    #featureVector.append(kf.ySymmetric2(ppImage))
    featureVector.append(kf.xSymmetric(ppImage))
```

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

```
    dataFeatures = []
    i=0
    while((i<images.shape[0])and(i<N)):
        dataFeatures.append(getFeatureVector(images[i]))
        i+=1
    return np.array(dataFeatures)


def getNeighbors(X, Z, xt, k, dist=euclideanDistance):
    distances = []
    for x1 in X:
        distance = dist(x1,xt)
        distances.append(distance)
    sortedDistances = distances[:]
    sortedDistances.sort()
    neighbors = []
    for d in sortedDistances:
        if(len(neighbors)<k):
            for index in range(len(distances)):
                if(d == distances[index] and len(neighbors)<k):
                    neighbor = (X[index],Z[index],d)
                    neighbors.append(neighbor)
    return neighbors
```

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

```
    return response

def getAccuracy(YT, ZT):
    success = 0
    size = len(YT)
    for i in range(size):
        if(YT[i] == ZT[i]):
            success+=1
    accuracy = (success/float(size))*100
    print(success)
    return accuracy

def predict(X, Z, XT, k):
    Y=[]
    i = 0
    for xt in XT:
        #print(datetime.datetime.now())
        #print(i)
        neighbors = getNeighbors(X, Z, xt, k)
        Y.append(getResponse(neighbors))
        i+=1
    return Y
```

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

```
def minuMean(dataFeatures):
    mean = meanVector(dataFeatures)
    dim = np.shape(dataFeatures)
    f = []
    for i in range(dim[0]):
        f.append(dataFeatures[i,:]-mean)
    return np.array(f)

def scaledFeaturesVectors(dataFeatures,eigVectors,eigValues,mean):
    sc = []
    dim = np.shape(dataFeatures)
    c = np.matmul(eigValues,np.transpose(eigVectors))
    for i in range(dim[0]):
        m = dataFeatures[i,:]-mean
        sc.append(np.matmul(c,m))
    return(np.array(sc))

def eigAnalysis(dataFeatures_train):
    mean = meanVector(dataFeatures_train)
    cov = np.cov(dataFeatures_train.T)
    eigValues,eigVectors = np.linalg.eig(cov)
    for i in range(eigValues.size):
        if eigValues[i]<0:
```

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

jupyter knn Last Checkpoint: Last Thursday at 9:11 AM (autosaved)
File Edit View Insert Cell Kernel Help Python 3

```
dataFeatures_train = getAllFeatureVectors(img_tr,Ntr)
eigValues,eigVectors,mean = eigAnalysis(dataFeatures_train)
dataFeatures_train = scaledFeaturesVectors(dataFeatures_train,eigVectors,eigValues,mean)

lbl_te, img_te = mylib.read("testing","../data")
dataFeatures_test = getAllFeatureVectors(img_te,Nte)
dataFeatures_test = scaledFeaturesVectors(dataFeatures_test,eigVectors,eigValues,mean)

k = 15
YT = predict(dataFeatures_train[0:Ntr], lbl_tr[0:Ntr], dataFeatures_test[0:Nte], k)
accuracy = getAccuracy(YT, lbl_te[0:Nte])
#print(np.array(YT))
#print(lbl_te[0:Nte])
print('Accuracy: ' + repr(accuracy))
print("end : "+ str(datetime.datetime.now()))
```

```
start : 2016-12-30 23:40:08.035360
(-4.87811954536e-15+4.18859484355e-15j)
(-4.87811954536e-15-4.18859484355e-15j)
8602
Accuracy: 86.02
end : 2016-12-31 00:43:53.497187
```

In [ ]: