

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [ ]: dfx=pd.read_csv('xdata.csv')
dfy=pd.read_csv('ydata.csv')
```

```
In [ ]: X=dfx.values
Y=dfy.values

X=X[:,1:]
Y=Y[:,1:].reshape((-1,))

print(X)
print(X.shape)
print(Y.shape)
```

```
In [ ]: plt.scatter(X[:,0],X[:,1],c=Y)
plt.show()
```

```
In [ ]: query_x=np.array([2,3])
plt.scatter(X[:,0],X[:,1],c=Y)
plt.scatter(query_x[0],query_x[1],color='blue')
plt.show()
```

```
In [ ]: def dist(x1,x2):
    return np.sqrt(sum((x1-x2)**2))

def knn(X,Y,queryPoint,k=5):

    vals=[]

    m=X.shape[0]

    for i in range(m):
        d=dist(queryPoint,X[i])
        vals.append((d,Y[i]))

    vals=sorted(vals)
    vals=vals[:k]

    vals=np.array(vals)

    # print(vals)

    new_vals=np.unique(vals[:,1],return_counts=True)
    print(new_vals)

    index=new_vals[1].argmax()
    pred=new_vals[0][index]

    return pred
```

```
In [ ]: knn(X,Y,query_x)
```

MNIST Datasets

```
In [ ]: df=pd.read_csv('train.csv')
print(df.shape)
```

```
In [ ]: print(df.columns)
```

```
In [ ]: df.head()
```

```
In [ ]: data=df.values
print(data.shape)
print(type(data))
```

```
In [ ]: X=data[:,1:]
Y=data[:,0]

print(X.shape,Y.shape)
```

```
In [ ]: split=int(0.8*X.shape[0])
print(split)
```

```
In [ ]: X_train=X[:split,:]
Y_train=Y[:split]
X_test=X[split:,:]
Y_test=Y[split:]

print(X_train.shape,Y_train.shape)
print(X_test.shape,Y_test.shape)
```

```
In [ ]: def drawImg(sample):
    img=sample.reshape((28,28))
    plt.imshow(img,cmap='gray')
    plt.show()
```

```
In [ ]: drawImg(X_train[3])
print(Y_train[3])
```

Prediction

```
In [ ]: pred=knn(X_train,Y_train,X_test[89])
print(pred)
```

```
In [ ]: drawImg(X_test[89])
print(Y_test[89])
```

```
In [ ]:
```