#importing libraries

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

#reading data

train=pd.read\_csv("E:\Datasets\MNIST\Train.csv")

submission=pd.read\_csv("E:\Datasets\MNIST\Test.csv")

y\_train=pd.read\_csv("E:\Datasets\MNIST\TrainLab.csv")

X\_train=train

X\_submission=submission

"""y\_train.head()

X\_train.head()"""

import matplotlib.colors as cm

normalize=cm.Normalize(vmax=1, vmin=0)

#PCA

from sklearn.decomposition import PCA

pca=PCA(n\_components=2)

pca\_result=pca.fit\_transform(X\_train)

print(pca.explained\_variance\_ratio\_)

print(X\_train.shape)

print(pca\_result.shape)

plt.scatter(pca\_result[:4000,0], pca\_result[:4000, 1],norm=normalize, edgecolor='none', alpha=0.5, cmap=plt.get\_cmap('bone', 10), s=5)

plt.colorbar()

pca=PCA(200)

pca\_full=pca.fit(X\_train)

plt.plot(np.cumsum(pca\_full.explained\_variance\_ratio\_))

plt.xlabel("# of components")

plt.ylabel("cumulative explained variance")