**HW2**

id:110062209

name:簡晟棋

**Function Implement**

**一張含有 文字, 螢幕擷取畫面, 軟體, 多媒體軟體 的圖片

自動產生的描述**sample\_by\_label:用於隨機抽取指定label的sample

i\_data為指定label(normal為0，anomaly為1)的所有sample

用np.random.choice隨機選擇min(number,len(i\_data))個sample的編號

(number可能超過len(i\_data))

**problem 1:**

一張含有 文字, 螢幕擷取畫面, 軟體, 字型 的圖片

自動產生的描述

用sample\_by\_label從train\_data抓出10個normal case，並從test\_data抓出10個anomaly case

用plt.plot作圖

­­**problem 2:**

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

沿用hw1的knn(return roc\_auc結果)

**problem 3:**

一張含有 文字, 螢幕擷取畫面, 軟體 的圖片

自動產生的描述

用train\_data去fit N個component的pca

對train\_data與test\_data做pca，結果為train\_data\_pca與test\_data\_pca

對train\_data\_pca與test\_data\_pca做pca.inverse\_transform，結果為train\_data\_r與test\_data\_r

算出test\_data與test\_data\_r的reconstruction error，以此算roc\_auc

用sample\_by\_label從train\_data\_r抓出10個normal case，並從test\_data\_r抓出10個anomaly case

用plt.plot作圖

return roc\_auc結果

**problem 4:**

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

對train\_data與test\_data做fft，結果為train\_data\_fft與test\_data\_fft

train\_data\_fft\_select與test\_data\_fft\_select是train\_data\_fft與test\_data\_fft每個case的lowest M frequency的項目

train\_data\_fft\_magnitude與test\_data\_fft\_magnitude是train\_data\_fft\_select與test\_data\_fft\_select的magnitude(絕對值)

將train\_data\_fft\_magnitude與test\_data\_fft\_magnitude用knn算roc\_auc\_score

先使train\_data\_ifft與test\_data\_ifft每個case的lowest M frequency的項目為train\_data\_fft與test\_data\_fft，後面的值補0，之後對train\_data\_ifft與test\_data\_ifft做fft

用sample\_by\_label從train\_data\_ifft抓出10個normal case，並從test\_data\_ifft抓出10個anomaly case

用plt.plot作圖

return knn的roc\_auc結果

**problem 5:**

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

haar:用遞迴式處理haar wavelet function

data只有兩項時，回傳(Alevel 1, Dlevel 1)

否則用dir == “right”算right = [Dlevel i]，用dir==”left”算left = [Alevel i]，把left遞迴下去後把遞迴結果跟right合併在一起後return

一張含有 文字, 螢幕擷取畫面, 軟體, 多媒體軟體 的圖片

自動產生的描述

level=np.ceil(np.log2(train\_data.shape[1]))為所需level數

L = int(2\*\*level)為所需長度

先使train\_data\_haar與test\_data\_haar為train\_data與test\_data每個case補0到所需長度L，接著用haar function 算出train\_data\_haar與test\_data\_haar經過haar wavelet function處理後的結果

將train\_data\_haar與test\_data\_haar用knn算roc\_auc\_score

return knn的roc\_auc結果

**Calculate, Record, Drawing**

**一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述**

用plt.show()把前面的plt部分畫出來

problem 1作圖結果

一張含有 文字, 圖表, 行, 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 文字, 繪圖, 圖表, 行 的圖片

自動產生的描述

problem 2結果:

Wafer:



ECG200:



problem 3 結果:

分別在N=1,2,5,10測試，發現Wafer效果在N=10效果最好，ECG200則在N=5效果最好

Wafer:

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述一張含有 文字, 圖表, 螢幕擷取畫面, 行 的圖片

自動產生的描述

ECG200:

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

一張含有 文字, 圖表, 繪圖, 行 的圖片

自動產生的描述

problem 4 結果:

分別在M=10,15,20,25測試，發現Wafer效果都差不多(M=25時比其他M值好一點點)，ECG200則在M=25效果最好

Wafer:

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述

一張含有 文字, 行, 圖表, 繪圖 的圖片

自動產生的描述

ECG200:

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

problem 5 結果:

分別在S=8,16,32,64測試，發現Wafer效果都差不多(S=8時比比其他S值好一點點)，ECG200則在S=32效果最好

Wafer:

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述

ECG200:

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述

**Observation**

**problem 6:**

在Wafer中，DWT(0.999)> DFT(0.998)>original knn(0.988)> PCA(0.974)

以下是可能原因:

Discrete Wavelet Transform:可以有效降維

Discrete Fourier Transform:資料週期性較佳

original : noise與維度太多

PCA:變異性差異不大

在ECG200中，PCA(0.948) DWT(0.948) > original knn(0.922) > DFT(0.857)

以下是可能原因:

PCA:可以有效地捕捉資料中的主要變異性

Discrete Wavelet Transform:可以有效降維

original : noise與維度太多

Discrete Fourier Transform:資料週期性較差

**problem 7:**

Wafer

original:

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

DFT:

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

DWT:

一張含有 文字, 螢幕擷取畫面, 字型, 印刷術 的圖片

自動產生的描述

best combination:

original : k=2

DFT : k=2, M=25

DWT : k=7, S=8

ECG200:

original:

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

DFT:

一張含有 文字, 螢幕擷取畫面, 字型, 設計 的圖片

自動產生的描述

DWT:

一張含有 文字, 螢幕擷取畫面, 字型, 印刷術 的圖片

自動產生的描述

best combination:

original : k=2

DFT : k=2, M=15

DWT : k=2, S=32