

# Digital Signal Processing Laboratory

## Lab 11 Environment Sound Classification

---

### 1. Introduction.

環境音的辨識能應用在許多地方，它能夠將有用的資訊從環境中萃取出來，進而達成辨識的效果，例如從吵雜的人車聲之中辨識出狗叫聲；亦或是從新輸入資料的一些特徵來比較出與環境音的不同，藉此偵測某些事件的發生，像是家裡從原本安靜的環境音變成小孩的哭聲而得知小孩的情況；再者，它也可以將被認為是環境音的聲音屏蔽掉，做出雜訊消除的功能，例如在講電話時自動將周遭雜音消除的技術。

### 2. Lab objectives.

此次實驗要我們對各種不同的環境音成分做分類，透過取 MFCC，將各種不同的聲音如狗吠聲、嬰兒哭聲、海浪聲、時鐘滴答聲、人類打呼聲等等進行特徵萃取，接著對 training data 做 cross validation，並對每個 training set 做 SVM(支援向量機)的分類，取 accuracy 最好的那個來當作最終的模型，最後用這個模型對 testing data 做預測。

### 3. Method.

一開始先 load 資料並對它們做整理、sorting 和 labeling，接著我透過 librosa 套件將每個檔案的音訊資料取出，用 MFCC 函數算出其特徵，其中我設定 MFCC 特徵的個數為 20，MFCC 的特徵數增加會讓辨識率增加，不過加到了一定程度後就沒甚麼影響了，而且反而會增加計算量與時間。

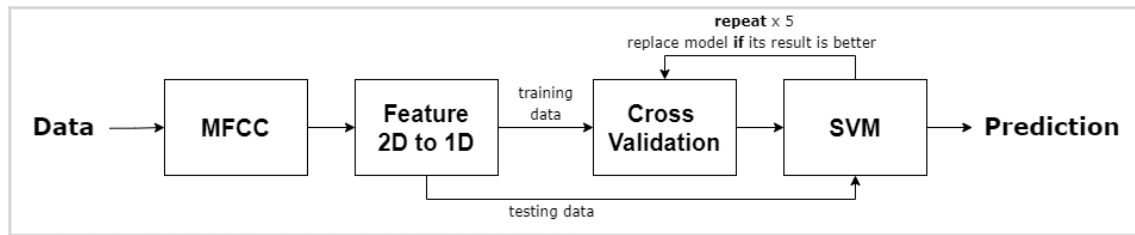
由於輸入進 SVM 的資料需要是一維的資料，但 MFCC 是二維的，因此我們需要將 MFCC 轉成一維的向量，為了達到此結果，我對每個 MFCC 特徵取平均與標準差，因此最終的 feature 數會有 40 個。

另外在將 MFCC 結果輸入 SVM 之前，我利用 KFold 這個函數做五次 cross validation，每次都會將原本的 training dataset 分為 training data 以及 testing data (validation)，並且會對兩者做 normalize 以降低資料中極端值對模型訓練的影響。

接著將資料輸入至 SVC 函數，SVC 函數是用來做 classification 的 SVM，有 kernel 和 C (Regularization parameter) 可以調整，kernel 有 linear、poly、rbf 和 sigmoid 這四種可以換，須根據資料的特性做最佳化；而 C 參數為懲罰係數，愈大代表錯誤的容忍程度愈低，在 training dataset 樣本中會區分的愈精細，因此設定太大容易造成 overfitting 的問題。反之若設定太小則會造成 underfitting。

經過各種嘗試後發現，SVM 對於此 dataset 而言，kernel 使用預設、C 設為 10 左右的時候，會有最好的 accuracy，而透過 confusion matrix 能可視化預測的結果。最後在確認模型的設定後，我拿原先已切好的 testing data 輸入進此模型中去進行預測，最終結果為 75%。

### 4. Flow chart.



## 5. Results and discussion of report questions.

### (1) Demo 1: Cross validation confusion matrix

No.	Confusion matrix	Accuracy																																																																																																																									
1	<div><p>Confusion matrix</p><table><tr><th>True label \ Predicted label</th><th>Dog bark</th><th>Rain</th><th>Sea waves</th><th>Baby cry</th><th>Clock tick</th><th>Person sneeze</th><th>Helicopter</th><th>Chainsaw</th><th>Rooster</th><th>Fire crackling</th></tr><tr><th>Dog bark</th><td>6</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Rain</th><td>0</td><td>9</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><th>Sea waves</th><td>0</td><td>0</td><td>2</td><td>0</td><td>2</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Baby cry</th><td>1</td><td>0</td><td>0</td><td>4</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr><tr><th>Clock tick</th><td>0</td><td>0</td><td>1</td><td>0</td><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Person sneeze</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>7</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Helicopter</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Chainsaw</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>4</td><td>0</td><td>1</td></tr><tr><th>Rooster</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td></tr><tr><th>Fire crackling</th><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>4</td></tr></table></div>	True label \ Predicted label	Dog bark	Rain	Sea waves	Baby cry	Clock tick	Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling	Dog bark	6	0	0	0	0	0	0	0	0	0	Rain	0	9	0	0	0	0	0	0	0	1	Sea waves	0	0	2	0	2	0	1	0	0	0	Baby cry	1	0	0	4	0	0	1	0	1	0	Clock tick	0	0	1	0	5	0	0	0	0	0	Person sneeze	0	0	0	0	0	7	0	0	0	0	Helicopter	0	0	0	0	0	0	6	0	0	0	Chainsaw	0	0	0	0	0	0	0	4	0	1	Rooster	0	0	0	0	0	0	0	0	2	0	Fire crackling	0	1	0	0	0	0	0	1	0	4	81.67 %
True label \ Predicted label	Dog bark	Rain	Sea waves	Baby cry	Clock tick	Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling																																																																																																																	
Dog bark	6	0	0	0	0	0	0	0	0	0																																																																																																																	
Rain	0	9	0	0	0	0	0	0	0	1																																																																																																																	
Sea waves	0	0	2	0	2	0	1	0	0	0																																																																																																																	
Baby cry	1	0	0	4	0	0	1	0	1	0																																																																																																																	
Clock tick	0	0	1	0	5	0	0	0	0	0																																																																																																																	
Person sneeze	0	0	0	0	0	7	0	0	0	0																																																																																																																	
Helicopter	0	0	0	0	0	0	6	0	0	0																																																																																																																	
Chainsaw	0	0	0	0	0	0	0	4	0	1																																																																																																																	
Rooster	0	0	0	0	0	0	0	0	2	0																																																																																																																	
Fire crackling	0	1	0	0	0	0	0	1	0	4																																																																																																																	
2	<div><p>Confusion matrix</p><table><tr><th>True label \ Predicted label</th><th>Dog bark</th><th>Rain</th><th>Sea waves</th><th>Baby cry</th><th>Clock tick</th><th>Person sneeze</th><th>Helicopter</th><th>Chainsaw</th><th>Rooster</th><th>Fire crackling</th></tr><tr><th>Dog bark</th><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Rain</th><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>2</td></tr><tr><th>Sea waves</th><td>1</td><td>0</td><td>3</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Baby cry</th><td>0</td><td>0</td><td>0</td><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Clock tick</th><td>0</td><td>0</td><td>0</td><td>0</td><td>7</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Person sneeze</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Helicopter</th><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Chainsaw</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>2</td></tr><tr><th>Rooster</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>7</td><td>0</td></tr><tr><th>Fire crackling</th><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>4</td></tr></table></div>	True label \ Predicted label	Dog bark	Rain	Sea waves	Baby cry	Clock tick	Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling	Dog bark	5	0	0	0	0	0	0	0	0	0	Rain	0	6	0	0	0	0	0	1	0	2	Sea waves	1	0	3	1	0	0	0	0	0	0	Baby cry	0	0	0	4	0	0	0	0	0	0	Clock tick	0	0	0	0	7	0	1	0	0	0	Person sneeze	0	0	0	0	0	4	0	0	0	0	Helicopter	0	0	0	2	0	0	6	0	0	0	Chainsaw	0	0	0	0	0	0	0	3	0	2	Rooster	0	0	0	0	0	0	0	0	7	0	Fire crackling	0	0	1	0	0	0	0	0	0	4	81.67 %
True label \ Predicted label	Dog bark	Rain	Sea waves	Baby cry	Clock tick	Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling																																																																																																																	
Dog bark	5	0	0	0	0	0	0	0	0	0																																																																																																																	
Rain	0	6	0	0	0	0	0	1	0	2																																																																																																																	
Sea waves	1	0	3	1	0	0	0	0	0	0																																																																																																																	
Baby cry	0	0	0	4	0	0	0	0	0	0																																																																																																																	
Clock tick	0	0	0	0	7	0	1	0	0	0																																																																																																																	
Person sneeze	0	0	0	0	0	4	0	0	0	0																																																																																																																	
Helicopter	0	0	0	2	0	0	6	0	0	0																																																																																																																	
Chainsaw	0	0	0	0	0	0	0	3	0	2																																																																																																																	
Rooster	0	0	0	0	0	0	0	0	7	0																																																																																																																	
Fire crackling	0	0	1	0	0	0	0	0	0	4																																																																																																																	
3	<div><p>Confusion matrix</p><table><tr><th>True label \ Predicted label</th><th>Dog bark</th><th>Rain</th><th>Sea waves</th><th>Baby cry</th><th>Clock tick</th><th>Person sneeze</th><th>Helicopter</th><th>Chainsaw</th><th>Rooster</th><th>Fire crackling</th></tr><tr><th>Dog bark</th><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Rain</th><td>0</td><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><th>Sea waves</th><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>2</td><td>0</td></tr><tr><th>Baby cry</th><td>0</td><td>0</td><td>0</td><td>8</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr><tr><th>Clock tick</th><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Person sneeze</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><th>Helicopter</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>5</td><td>0</td><td>1</td><td>0</td></tr><tr><th>Chainsaw</th><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>5</td><td>0</td><td>1</td></tr><tr><th>Rooster</th><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>4</td><td>0</td></tr><tr><th>Fire crackling</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>5</td></tr></table></div>	True label \ Predicted label	Dog bark	Rain	Sea waves	Baby cry	Clock tick	Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling	Dog bark	4	0	0	0	0	0	0	0	0	0	Rain	0	4	0	0	0	0	0	0	0	1	Sea waves	0	0	1	0	1	0	1	0	2	0	Baby cry	0	0	0	8	0	0	1	0	1	0	Clock tick	0	0	0	0	3	0	0	0	0	0	Person sneeze	0	0	0	0	0	6	0	0	0	0	Helicopter	0	0	0	0	0	0	5	0	1	0	Chainsaw	0	0	0	0	1	1	0	5	0	1	Rooster	0	0	2	0	0	0	0	0	4	0	Fire crackling	0	0	0	0	0	0	1	1	0	5	75.00 %
True label \ Predicted label	Dog bark	Rain	Sea waves	Baby cry	Clock tick	Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling																																																																																																																	
Dog bark	4	0	0	0	0	0	0	0	0	0																																																																																																																	
Rain	0	4	0	0	0	0	0	0	0	1																																																																																																																	
Sea waves	0	0	1	0	1	0	1	0	2	0																																																																																																																	
Baby cry	0	0	0	8	0	0	1	0	1	0																																																																																																																	
Clock tick	0	0	0	0	3	0	0	0	0	0																																																																																																																	
Person sneeze	0	0	0	0	0	6	0	0	0	0																																																																																																																	
Helicopter	0	0	0	0	0	0	5	0	1	0																																																																																																																	
Chainsaw	0	0	0	0	1	1	0	5	0	1																																																																																																																	
Rooster	0	0	2	0	0	0	0	0	4	0																																																																																																																	
Fire crackling	0	0	0	0	0	0	1	1	0	5																																																																																																																	

4	<p>Confusion matrix</p> <p>True label</p> <p>Predicted label</p>	80.00 %
5	<p>Confusion matrix</p> <p>True label</p> <p>Predicted label</p>	81.67 %

## (2) Demo 2: Testing evaluation confusion matrix

Testing evaluation	<p>Confusion matrix</p> <p>True label</p> <p>Predicted label</p>	Accuracy = 75.0 %
-----------------------	--	----------------------

## 6. Conclusion.

這次實驗利用 MFCC 取特徵、再透過 cross validation 對 SVM 模型做最佳化，讓我們實作出各種聲音的辨識，也讓我對於音訊處理的流程有了更深的認識，另外我也了解到模型參數與訓練資料集對於模型的辨識率的影響，有時得需要透過不斷嘗試，才能調出最好的預測結果，然而，這又不代表模型對於新的沒看過的測試資料有著最好的辨識率，因為在訓練模型時為了使結果最好，常常會不小心讓模型對於訓練資料的過擬合。另外，在取 feature 時也很重要，像是我同學用的所有模型參數都跟我一樣，但是他在 feature 中多加了 zero crossing 的特徵，就讓他的 testing data 的預測結果達到 80%，比我整整高出了 5%，可見在整個音訊辨識的處理流程中，每個步驟都會對預測結果造成不小的影響。