A1105505 林彧頎

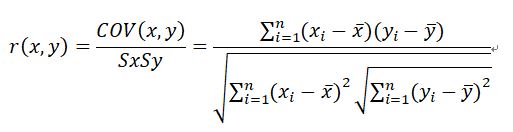
皮爾森相關分析用於探討兩連續變數(X, Y)之間的**線性相關**(所以不能做非線性的分析，如果要做非線性的分析需要用其他的相關性分析法)，若兩變數之間的相關係數絕對值較大，則表示彼此相互共變的程度較大。一般而言，若兩變數之間為正相關，則當X提升時，Y也會隨之提升；反之，若兩變數之間為負相關，則當X提升時，Y會隨之下降。

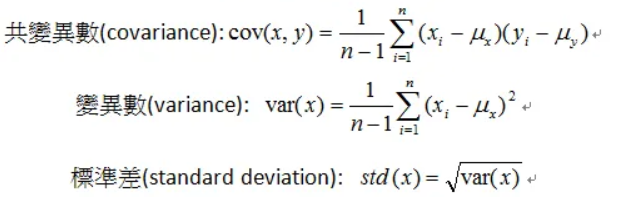
其中分母是在做正規化，分子則是COV(共變數)，r為相關係數就是老師ppt裡的ρ，會落在-1到1之間: -1≤ρ≤1，μx和μy 分別代表變數x和y的平均數。

共變異數衡量兩個隨機變量之間的線性關係，即它們的變動趨勢是否一致。

變異數是變異數衡量一個隨機變量在其平均值周圍的數據分散程度。

**皮爾森相關係數公式為：(網路上的公式(同PPT))**

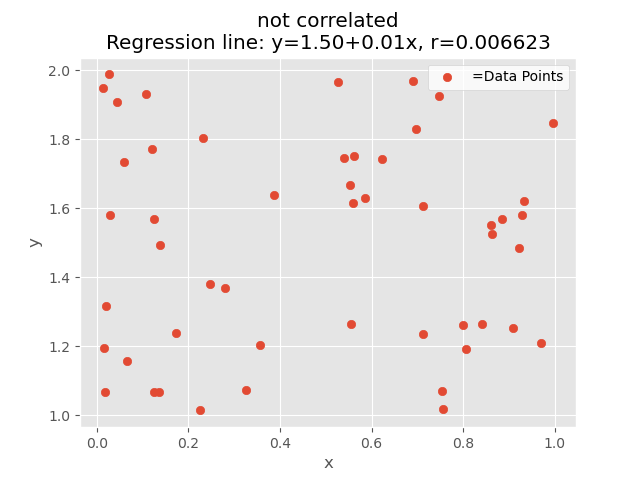


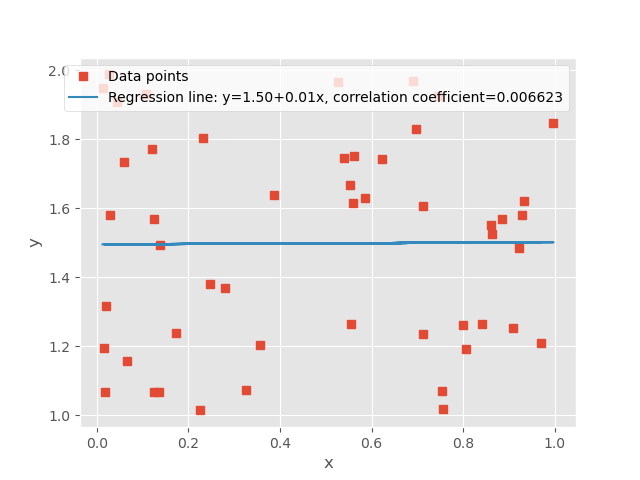


我學習到可以利用此方法來先對數據分析前進行相關係數的比較，財部會使模型訓練的結果很低，使在距相關性的數據下進行分析達到更好的效果。

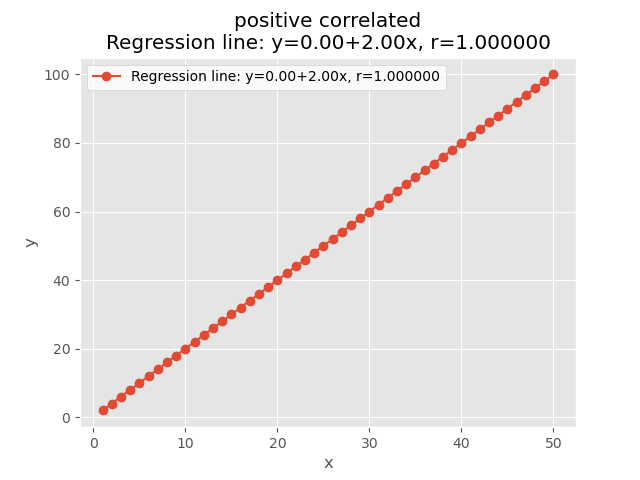
(下一頁說明PPT五個範例)

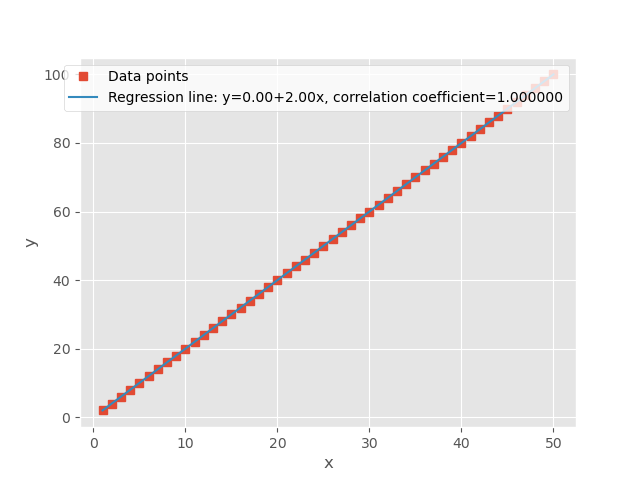
(程式碼與圖片位置請記得看ReadMe)

1. 第一筆不具相關性，首先它的相關係數很低，表示x與y之間變動趨勢不一致，而且數據分散程度高，第一張圖是照課本的，第二張圖是量化的線性指標。

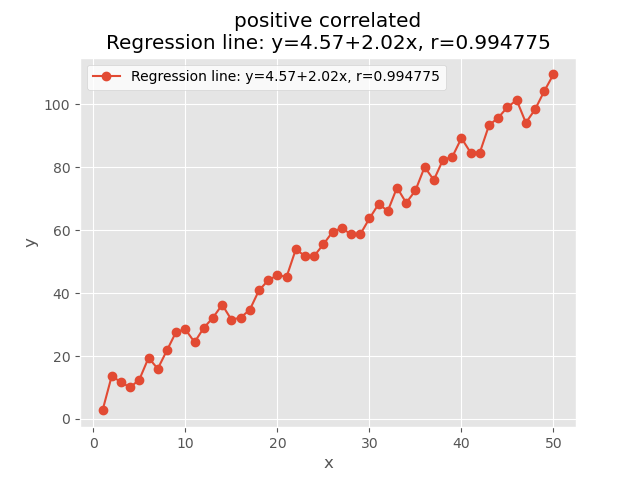


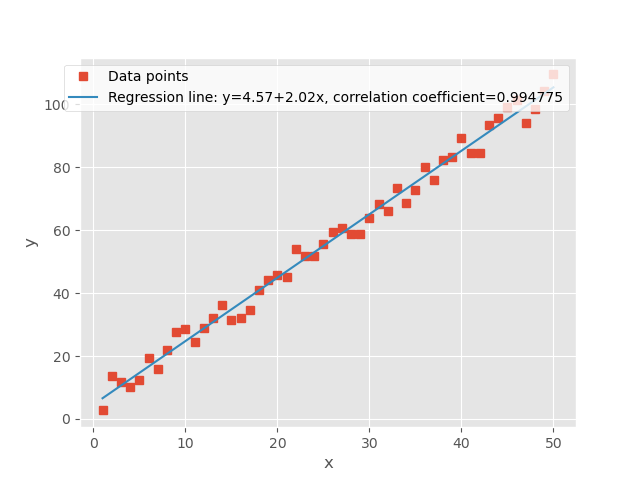
2. 第二筆具正相關性，由於他的x與y之間變動趨勢十分一致，而且數據分散程度低，導致可以看到他的相關係數絕對值為最高值1，第一張圖是照課本的，第二張圖是量化的線性指標。



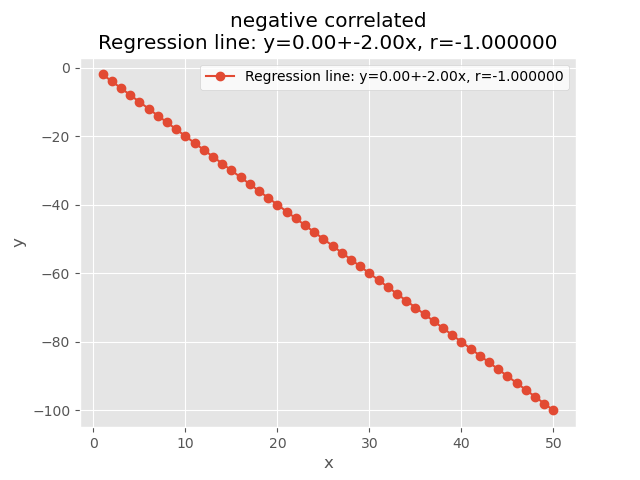


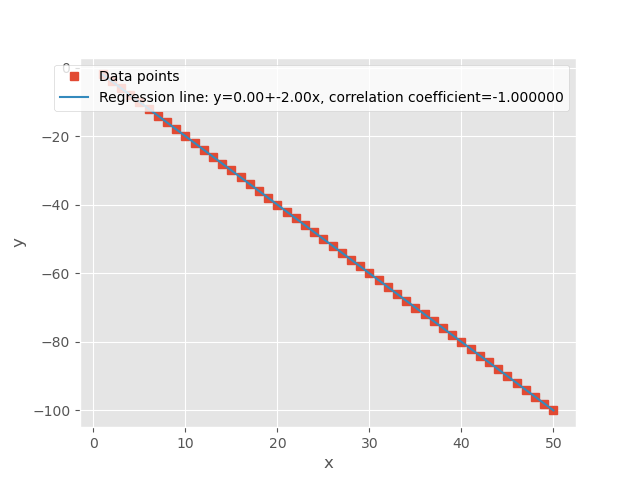
3. 第三筆具正相關性，由於他的x與y之間變動趨勢十分一致，而且數據分散程度低，導致可以看到他的相關係數絕對值為近乎最高值，然而這裡的相關係數值與PPT不同，老師說這是正常的因為數據可能有一點點不一樣或是因模型有些不同，第一張圖是照課本的，第二張圖是量化的線性指標。





4. 第四筆具負相關性，由於他的x與y之間變動趨勢十分一致，而且數據分散程度低，導致可以看到他的相關係數絕對值為最高值1(-1)，第一張圖是照課本的，第二張圖是量化的線性指標。





5. 第五筆具負相關性，由於他的x與y之間變動趨勢十分一致，而且數據分散程度低，導致可以看到他的相關係數絕對值為近乎最高值，第一張圖是照課本的，第二張圖是量化的線性指標。

