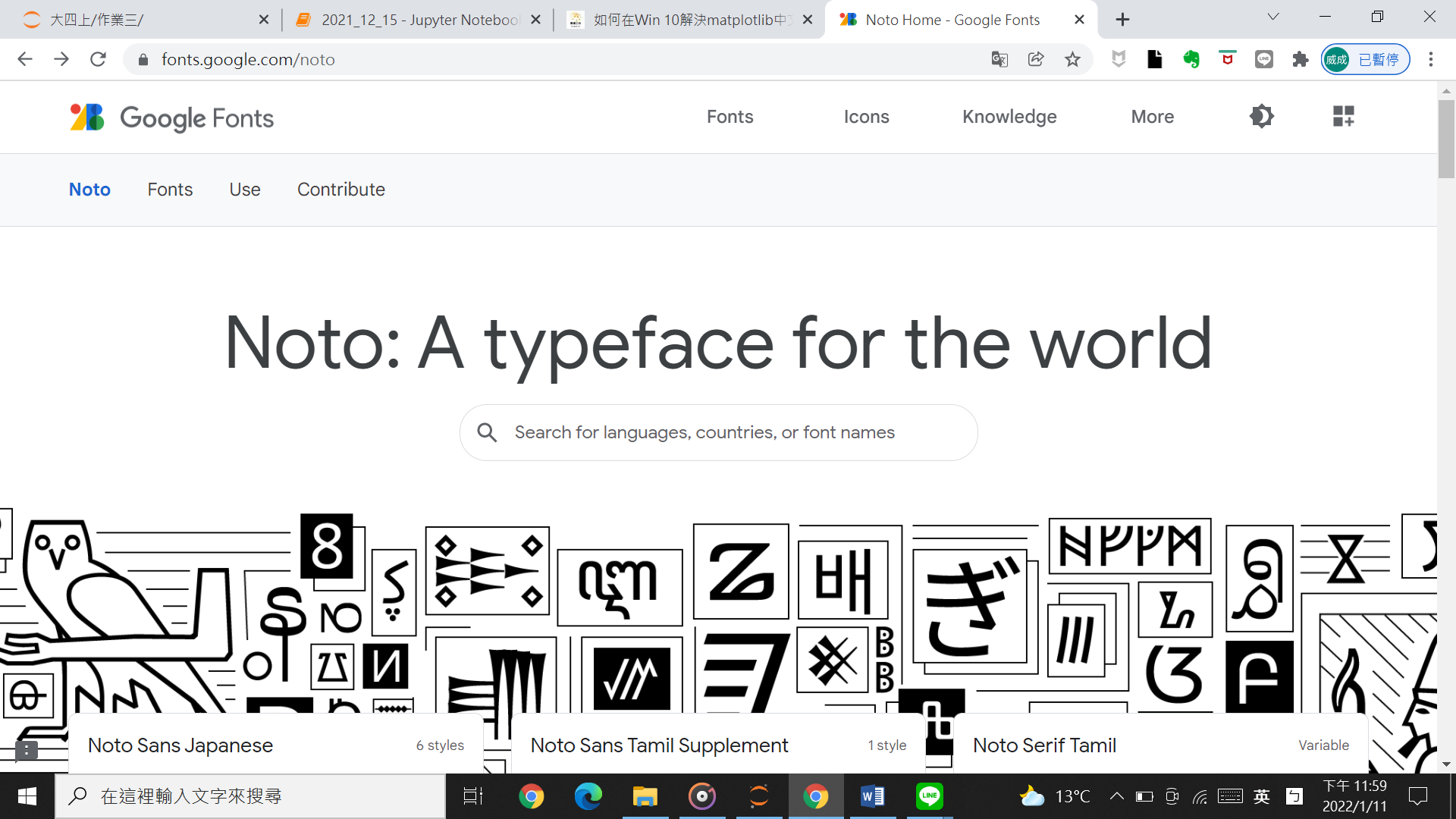
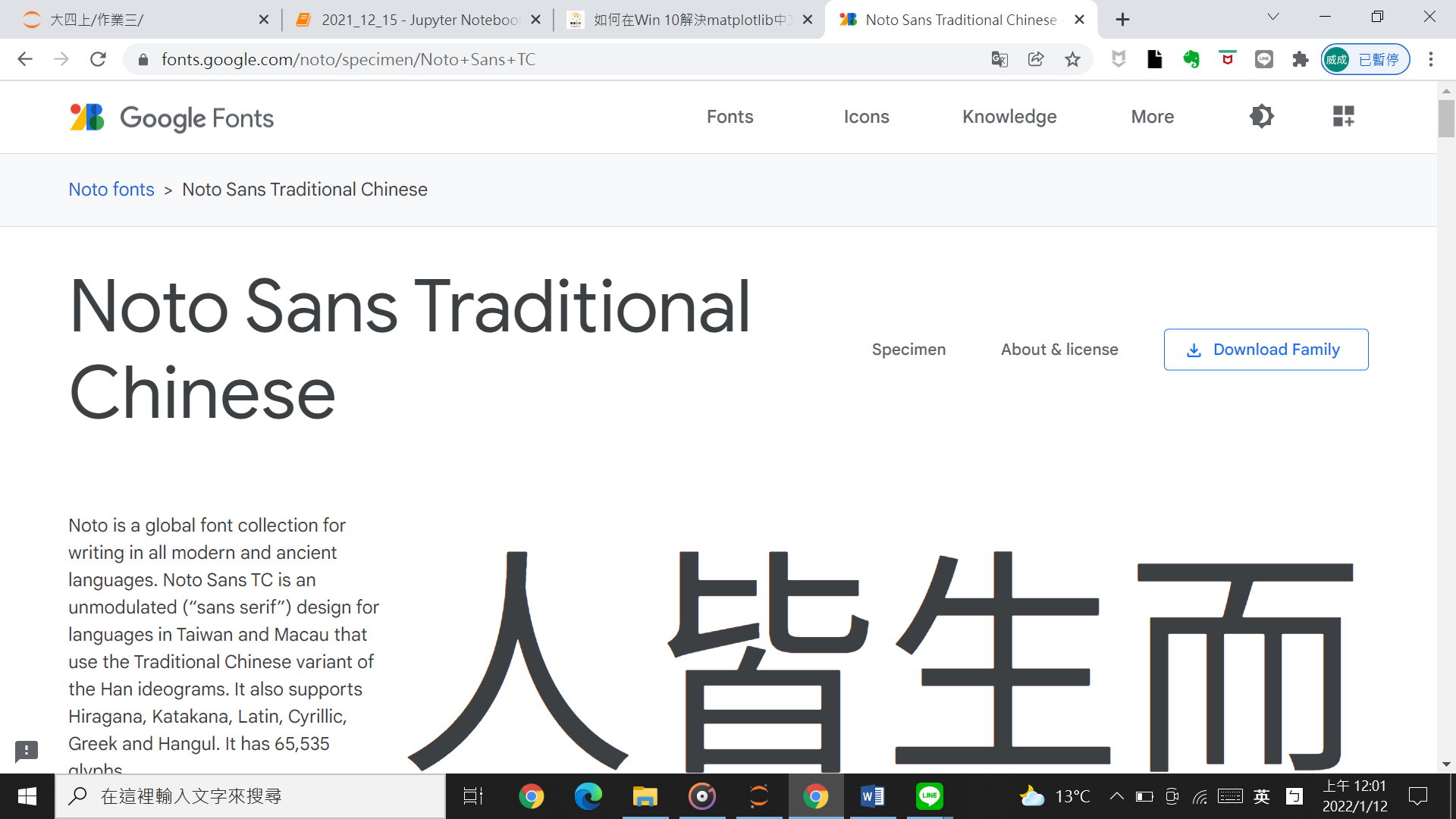
大數據分析技法 作業三

1. 目標
   1. 顯示出三家公司股票股價趨勢
   2. 做出關於股票資料的黃金交叉點
   3. 利用平均值、標準差、偏度和峰度的概念，分析股票資料分布的狀態
   4. 套入投資組合報酬率、權重、貝氏定理和投資風險的概念，分析股票資料
   5. 利用陳傲賢所撰寫相關深度學習的程式，來預測未來幾天投資組合報酬率，並且把貝氏定理和投資風險納入考量，輸入的資料為每家公司股票的權重
2. 完成項目
   1. 顯示出三家公司股票股價趨勢
   2. 做出關於股票資料的黃金交叉點
3. 遭遇困難和如何解決
   1. 我想要把其中一家公司的股票股價趨勢以整個年度的方式呈現出來，y軸放股價，x軸就放1月、2月…到12月，但是那時候遇到的困難是中文沒辦法顯示出來，會以亂碼的方式呈現。不過，後來找到的方法就是到Google Fonts網頁下載中文，如下圖所示

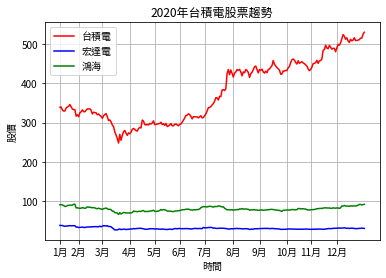


我是下載Noto Sans TC，中文版，詳細資訊可以參考6.1的連結，如下圖所示

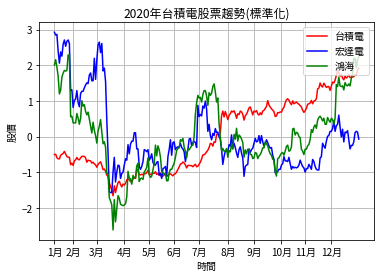


下載完Noto Sans TC，我在Jupyter Notebook那邊打上print(matplotlib.\_\_file\_\_)，就可以知道Noto Sans TC放在哪個路徑下， Noto Sans TC的otf檔要放在D:\Users\使用者\anaconda3\envs\geo\_env\Lib\site-packages\matplotlib\mpl-data\fonts\ttf中。關於這個問題其實我卡蠻久的，那時候我在Jupyter Notebook額外新增一個環境，而且我在新增的環境中執行程式碼，然後沒有意識到要在Jupyter Notebook打上print(matplotlib.\_\_file\_\_)，讓我以為Noto Sans TC要放在D:\Users\使用者\anaconda3\Lib\site-packages\matplotlib\mpl-data\fonts\ttf，然後不管用哪種方式，最後顯示出來的圖都沒辦法顯示出中文，光這一點就卡很久，不過，知道放錯資料夾之後，就可以正常顯示中文了。加入完Noto Sans TC之後，刪掉.matplotlib檔，然後在Jupyter Notebook打上plt.rcParams[‘font.sans-serif’]=[‘Noto Sans TC’]，重新執行程式碼，問題就解決了。詳細資訊可以參考6.2連結。

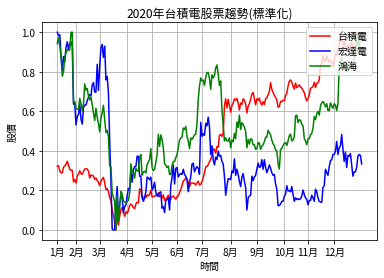
* 1. 如下圖所示，我想要分析三家公司的整個年度股價的趨勢，不過，我遇到一個問題就是台積電的股價和宏達電以及鴻海差距太大了，雖然可以知道台積電整個年度股價的變化，但是我看不出來宏達電和鴻海股價的變化，所以沒辦法分析這三家公司股票的資料



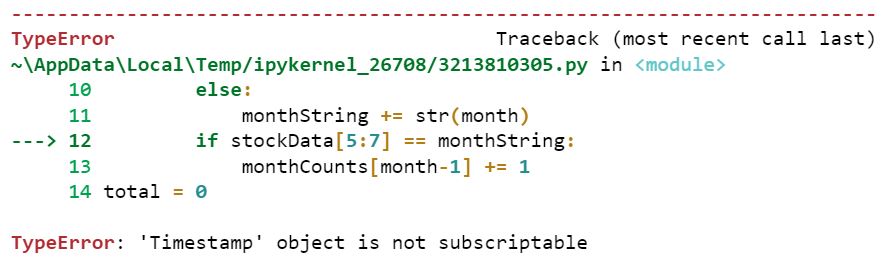
後來我想到的一個方法就是把這三家公司股票的資料給標準化，這樣我就可以知道這三家公司股價整個年度的趨勢，如下圖所示，我先用Z分數標準化的方式把所有資料給標準化，把所有股價濃縮到-3到3，從這張圖可以知道台積電在2020年中，整個股價還是有往上升的現象，而宏達電，整個股價呈往下跌的狀態，尤其是3月到4月期間，跌的幅度非常的高，從這邊也可以知道，我猜應該蠻多人會非常想要賣出宏達電股票。另外，鴻海整個2020年，股價我感覺就是有下跌也有上漲，不過上漲或下跌的幅度蠻高的，因此我猜買鴻海股票的人應該都蠻恐慌的，因為鴻海股票整個股價的波動非常地大。



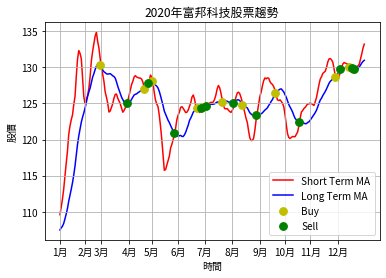
另外，我還有用另一種方法把資料給標準化，方法為Max-Min標準化，如下圖所示，跟Z分數標準化的差別就是，把所有股票股價的數據濃縮到0到1之間，三家公司的股票股價的趨勢都一樣，不過，Z分數標準化出來的資料，宏達電公司的線和鴻海的線一路上幾乎都貼在一起。而Max-Min標準化出來的結果，宏達電公司的線和鴻海的線就分開來了，尤其是在後面就會很明顯，原因為什麼會這樣，我再猜應該是算法上導致出來的結果，事實上，真正的原因我沒有花時間去研究，所以目前還不知道為何有這樣子的現象發生。



* 1. 遇到TypeError的問題，後來發生理由為資料是datetime型態，解決方法就是使用他給的方法，可以參考6.11連結



* 1. 紅色的線為5天股價的平均，藍色的線為20天股價的平均，當紅線超過藍線時，就會標註綠點，也就是買，相反地，當紅線跑到藍線以下時，就會標註黃點，也就是賣，目前遇到一個問題就是，所標註的黃點和綠點有時候不會在兩條線的交叉點上，也不知道如何做會更精準的標註在交叉點上，我目前的做法就是把紅線和藍線前後兩天的股價做平均，並且當作是標註的點



黃色改成Sell，綠色改成Buy

1. 程式碼
   1. <https://drive.google.com/drive/folders/14VmZ_2Rzp7G4O90XbHABf-MirKpKusK2?usp=sharing>

點擊此連結可以看到程式碼和資料集

1. 參考來源
   1. [Noto Home – Google Fonts](https://fonts.google.com/noto)
   2. [如何在Win10解決matplotlib中文顯示的問題](https://daxpowerbi.com/%E5%A6%82%E4%BD%95%E5%9C%A8win-10%E8%A7%A3%E6%B1%BAmatplotlib%E4%B8%AD%E6%96%87%E9%A1%AF%E7%A4%BA%E7%9A%84%E5%95%8F%E9%A1%8C/)
   3. [Preprocessing data](https://scikit-learn.org/stable/modules/preprocessing.html)
   4. [【資料科學】-資料的正規化與標準化](https://aifreeblog.herokuapp.com/posts/54/data_science_203/)
   5. [Python資料預處理: 資料標準化](https://www.itread01.com/content/1543176304.html)
   6. [sklearn.preprocessing.MinMaxScaler](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html)
   7. [黃金交叉指標算法【Python量化交易】](https://zhuanlan.zhihu.com/p/371079405)
   8. [Download Financial Dataset Using Yahoo Finance in Python | A Complete Guide](https://www.analyticsvidhya.com/blog/2021/06/download-financial-dataset-using-yahoo-finance-in-python-a-complete-guide/)
   9. [yfinance攻略!Python下載股票價格數據無難度](https://pythonviz.com/finance/yfinance-download-stock-data/)
   10. [使用Python及Yahoo Finance API抓取台股歷史資料](https://aronhack.com/zh/retrieve-stock-historical-data-with-python-and-yahoo-finance-api/)
   11. [Time series / date functionality](https://pandas.pydata.org/pandas-docs/version/0.25.3/user_guide/timeseries.html)