機器學習實務與應用

Homework #2
Due 2019 March 4 9:00AM

1. 《網路爬蟲》撰寫 python 程式自動至蘋果日報(https://tw.appledaily.com)截取其即時新聞之時間及完整標題。如底下結果：


```
22:01 打壓土包子？韓國瑜等嘸吉隆坡市長　議員早餐會也破局
21:58 【HBL】松山止於8強3連霸夢碎　黃萬隆向球迷說明年見
21:56 【二次川金會】終於定案！川普明晚到越南　27日下午雙方會面
21:55 陳柏融自費回台跑宣傳　遇學生妹憶高中糗事
21:52 謝淑薇最新世界排名第27　重返6年新高
21:49 黑幫闖民宅開槍尋仇　歹勢！勇媽嗆這句讓對方認錯撤退
21:47 酒駕重罰修法　綠委：待法務部版本提出後排審
21:38 簽MOU對象有統戰背景　基進黨批韓國瑜賣台
```

為取得完整標題必須要進入各個標題連結的網頁。底下敘述相關提示：

- 可以透過 requests.get()取得蘋果日報首頁之 html
- 將取得之 html 物件透過 BeautifulSoup(html.text,'html.parser')針對 html 標簽進行解釋和分類。為了能從 html 中抓取目標的物件，必須對 html 格式有初步的了解
- 蘋果日報首頁之 html 中的即時新聞的放置於以標籤<aside>標示的區域元素。可以透過 BeautifulSoup 之 find('aside')找出
- 再透過 find('time')進一步找出時間資訊
- 透過 find('a') 可以找到各個標題連結之網址
- 連結入各個網址找出標題

上述若有不清楚之處，可以從網路找尋相關資訊。

請將找出的時間與標題，計算每個標題的字數。之後再將時間、標題與字數透過 pandas 套件建立一個 dataframe。最後存成 headline.xlsx 檔。

2. 《機率》利用 python 實作底下三題機率問題。限用 numpy、matplotlib、math、random 等基本函式庫，但禁止使用 numpy.random.binomial 與 numpy.random.multinomial 或其他機率模型相關的函式庫。

甲、擲硬幣的狀況為一種白努利機率的結果，假設擲正面定義為 1，擲反面定義為 0，如果此硬幣為不均勻的硬幣，擲正面的機率 p 為 0.7，擲此硬幣 1000 次，那正面共有幾次?反面會有幾次?結果用 matplotlib 將直方圖畫出來

乙、呈上題，若將此不均勻硬幣”連續擲 5 次”，結果可能會有五反、4 反 1 正、3 反 2 正、2 反 3 正、1 反 4 正、5 正，將這五種狀況分別定義為 0、1、2、3、4，若將”連續擲五次硬幣”此狀況執行 1000 次(此分布結果為二項分布 binomial 結果)，將每種狀況對應的次數有幾次?次數分布的用直方圖畫出來。

丙、若有一枚不均勻的六面骰，1~6 的結果機率分別為 0.05、0.10、0.15、0.20、0.25、0.25，此機率結果為 mulitnoulli，若擲這顆骰子 1000 次，試問擲到 1~6 的結果分別有幾次?，將分布結果的直方圖畫出來。