

機器學習導論

Homework #1

Due 2022 Sep 19 11:00PM

1. **說明**：請將附件 *image.rar* 解壓縮於所撰寫 python 程式之工作目錄下。並撰寫具有以下功能之 python 程式：

- 可藉由 `input()` 輸入一個 1 到 20 的數字 *n*，如下圖。若輸入格式有誤，或非在 1 到 20 之間，都應跳出錯誤訊息，要求重新輸入。譬如輸入 -2, 3.4, aa, -7 或是 21, 35 都將視為輸入錯誤。

請選擇想顯示之圖片編號(1-20):

- 之後依據輸入數字 *n* 從放置於目錄 `./test` 下之檔案 *picn.jpg* 的圖片顯示於螢幕上。譬如：若輸入 15，則 *pic15.jpg* 將顯示於螢幕，如下：



- `./reference` 目錄下的子目錄 `cats/` 及 `dogs/` 已分別收錄 50 張貓及狗的圖片。請從這 100 張圖片中，找出最接近所選擇圖片之 3 張圖片，並將結果顯示於螢幕上。判斷兩張圖片相似度的指標可以採用底下公式：

$$\frac{1}{H \times W \times 3} \sum_{c=0}^2 \sum_{i=0}^H \sum_{j=0}^W |I_1(i, j, c) - I_2(i, j, c)|$$

其中圖片解析度 $H \times W$ ， c 則代表像素的 R、G 或 B 不同通道的成分， $I_1(i, j, c)$ 代表圖片 I_1 轉成 ndarray 對應第 i 列、 j 行像素之 c 通道強度。*由於圖片解析度可能不同，請先將之統一 **resize 到 64x64** 再做比較。* (請注意在計算上式圖片差異過程中，是否會有 `overflow` 的情況發生)

- 請依此判斷該圖片應為狗或者是貓。
- 也請於 `jupyter notebook` 上面討論，此一判斷所選擇之圖片是貓還是狗的方式可能的問題。

繳交說明：請繳交 `jupyter notebook` 之檔案。其中 *image.rar* 可以不用再附上

2. **說明**：撰寫一個 python 程式利用 `pandas` 讀取 *HW1_data.csv* 資料：

- 首先轉換成如下圖的 `dataframe` 的格式

| | Name | Sex | Age | HR | Height | Weight | BP |
|---|--------|-----|-----|----|--------|--------|-----|
| 0 | Emma | F | 35 | 61 | 159 | 64 | 137 |
| 1 | Olivia | F | 51 | 92 | 152 | 52 | 100 |
| 2 | Mason | M | 37 | | 186 | 76 | 134 |

- 計算出 Age, HR, Height, Weight 及 BP 的平均值加於 dataframe 最下列。(請注意避免累加過程中出現 overflow 現象)
- 利用程式找出並於螢幕列出各分項指標 feature (Age, HR, Height, Weight 及 BP) 中之最大者的姓名 (Name)。
- 請繪製身高體重 (Height, Weight) 之散佈圖 (Scatter Plot)，女生請以紅點標示，男生請以藍點標示。
- 繪製不同年齡 (Age) 區間的人數直方圖及圓餅圖 (Pie Chart)。請依 1-10, 11-20, 21-30...每 10 歲為一個區間統計。
- 繪製男女性別分佈比例之圓餅圖 (Pie Chart)。

繳交：請繳交 jupyter notebook 之檔案。