機器學習實務與應用

Homework #3
Due 2019 Mar 11 9:00AM

1. 利用 numpy.random 可以產生各種許多常用分佈的隨機資料，也可以從一個一維數字陣列中隨機抽取一定長度的樣本。撰寫 python 程式去驗證中央極限定理，也就是從服從任何分佈的母體中，當抽樣樣本愈大，抽樣樣本的平均數分佈將接近 normal distribution。可以使用（但不限用）seaborn 套件去繪製分佈圖。
   (若要使用 seaborn 套件中 kdeplot 的功能，記得將電腦安裝的 statsmodel 升級至 0.8.0 版)

2. 附檔 weight-height.csv 共有一萬筆包含性別、身高及體重的資料。
   i. 請利用附檔資料進行身高的抽樣，樣本的抽樣數分別為 n=2、10、30、100，每種樣本抽樣皆重複進行 1000 次。將其平均值以值方圖的方式呈現。請於 jupyter notebook 中附加討論樣本的抽樣數變大，所得到的隨機樣本分佈的情況
   ii. 請由檔案隨機抽樣男女性別各 50 筆資料，使用 statsmodels 套件，利用 t-test 計算出相對應的 P 值，討論「性別因素是否影響身高」。
   iii. 請計算出附檔「身高—體重」的相關係數。

3.