HOMEWORK 1 成大化學系大三 學號 C34066014 洪加城

1. 從Kaggle中找一個資料集, 描述此資料相關資訊, 包含: 問題定義,潛在問題,分析與預測難度及價值。

資料集Predict Molecular Properties

1. 問題定義-資料與問題同時

我們利用PubChem database的化學分子座標及其原子，和分子特性去擬合

1. 潛在問題

因為給定的是純Cartesian coordinate ,希望在空間旋轉且相對座標不變下得到相同數值，要有一個標準化過程讓同一個分子能被分辨出來。

每一個分子中，各鍵能位能圖並不會和標準原子一樣，隨著化學環境不同，共振系統等其他變因，使預測變得困難。

1. 分析與預測難度

單純直接用xyz座標下去做，我認為一定做不出一個好結果，我認為應該先轉成一些具化學意義的特徵在下去分析，像是轉成成分子內座標，標定出共軛範圍，定義出推拉電子基。要做出有化學意義的模型其實非常困難，首要如果能先預測出能量就已經是一大進展了。

1. 價值

現在的量子化學計算在電腦的快速發展已經逐漸可以運用在真實系統輔助證明機理，不過其中計算所的時間實在太長了，如果可以先建出一個小分子模型，並從這個慢慢建出高分子基團，對整個計算化學界可以說是一大翻盤，藥物發展,材料掃描或是有機合成都會有大大的進展。

1. 描述一個在你學習領域的資料或是日常生活中可取得的資料, 說明可以用來產生何種價值

其實我想說的已經在(a)提得差不多了，來另外提一個比較日常的，分析過去10年的GRE英文考試，用文章和問題標定出重要單字和文法，集結出一份大考精華，畢竟紅寶書有三萬多字。

這份精華可以快速幫助考生進入備考狀態，能夠大致上看得懂文章並快速釐清重點。