

HW-6

學號：R06521601 系級：土木所 姓名：黃伯凱

(1%) 1. 請比較有無 `normalize` 的差別。並說明如何 `normalize`。

答：就像過去作業一和作業二一樣，看到這種類型的 `training` 就會很直覺的要做 `normalize`，這次我將 `training` 透過 `std` 的方式做 `scaling`，有 `normalize` 的精度比沒有的高了大約 0.009 多，這次的落差沒有作業一和作業二那麼多，我想主要是數值的波動沒那麼大，不像作業二有差超級多的狀況。

(1%) 2. 比較不同的 `embedding dimension` 的結果。

答：原本我是使用的是 64 Dimension，另外多做了 128 以及 256 Dimension，上船咖狗之後發現好像沒有差太多，分別是 0.86053、0.86066、0.86044。

(1%) 3. 比較有無 `bias` 的結果。

答：這題在作業一的時候好像有請大家思考過，那時候就發現其實 `bias` 能夠幫助我們校正精準度，有時候雖然訓練的結果恰似已經到了最低點，但其實可能是集中在某處，需要一個 `bias` 來協助校正到最低點。而且就算這個未知的權重其實沒有 `bias`，但訓練的結果理應將我們加入的 `bias` 趨近於 0。

(1%) 4. 請試著將 `movie` 的 `embedding` 用 `tsne` 降維後，將 `movie category` 當作 `label` 來作圖。

答：降維後畫出來的圖如下，淺紫色的代表 `Children's|Animation`，可以看的出來比較集中在中間右方，藍色 `Adventure|Western|Comedy` 最複雜，幾乎遍布了整張圖片，橘色為 `Thriller|Horror|Mystery` 則分布在中間下側居多。



(1%)5. 試著使用除了 **rating** 以外的 **feature**, 並說明你的作法和結果, 結果好壞不會影響評分。

答：一開我真心覺得年紀應該會是看電影的一個重要依據才對呀，但後來測試的結果拿掉年紀之後竟然提升不少，我試著去看 **User** 的資料集，發現有些使用者年齡似乎太過不真實，有不少筆資料顯示使用者只有一歲，冏。在我還沒挑掉年紀這項 **feature** 時，精度大概落在 0.875 多，一直降不下去，挑掉之後就提升了不少，我發現除了年紀是個 **noise** 之外，可能也攸關我沒有對 **age** 做 **normalize**。