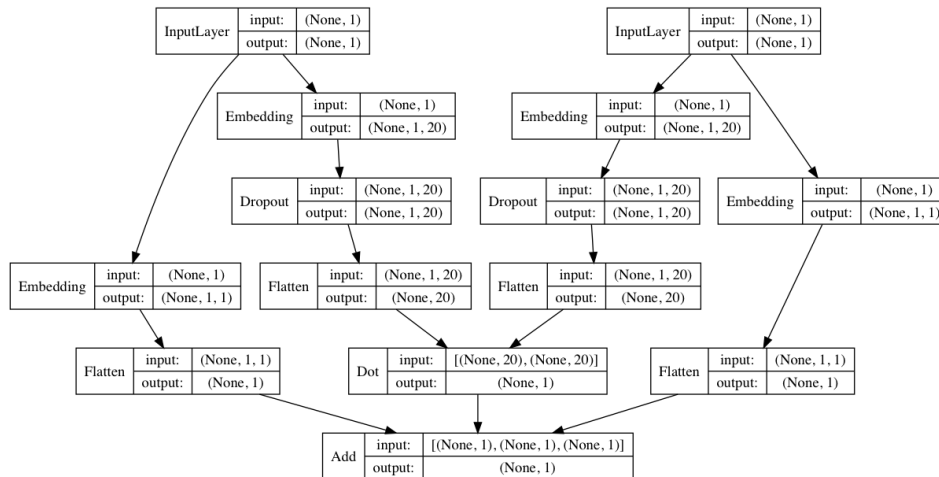


HW6

學號：B05902043 系級：資工二 姓名：劉鴻慶

先附上我的 MF model 架構，model 是參考去年的助教 Ppt 再加上 Dropout 層防止 overfitting，以下 report 前三題都是採用這個模型架構來做調整，比如說調整模型的 normalization, embedding dimension, bias 等。



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

Normalization method:

Training:

將每部電影的 Rating 減掉平均除上標準差，也就是 zero – mean

normalization : $R = (R - \text{mean}) / \text{std}$

Testing:

將預測出來的標準化分數還原成原始分數，才能取得預測的 Rating：

$\text{predict} = \text{std} * \text{predict} + \text{mean}$

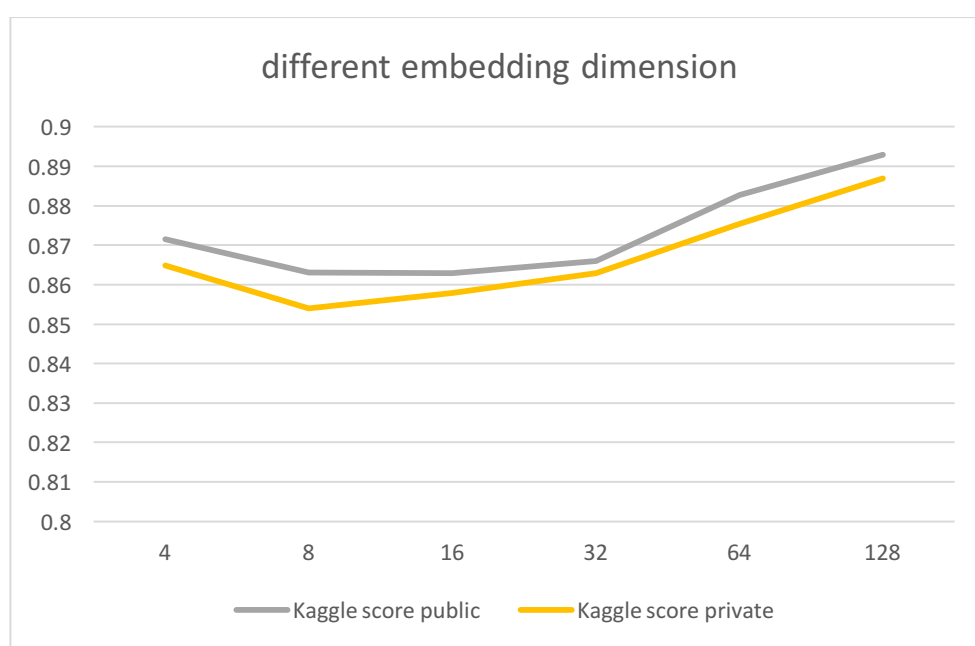
結果：

由結果可以發現有無 normalization 在 kaggle score 上的差異並不大

	沒有 normalization	有 normalization
Kaggle score public/private	0.86028/0.85371	0.86661/0.85791

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

dim	Validation loss	Kaggle score (public/private)
4	0.6067	0.87149/0.86482
8	0.5814	0.86314/0.85394
16	0.5795	0.86291/0.85788
32	0.6171	0.86598/0.86298
64	0.6212	0.88267/0.87541
128	0.6541	0.89295/0.88694



由上表可以發現 embedding dimension 太大或太小都會影響 model 的準確率，embedding dimension 要適中才可以得到較好的預測結果。

3. (1 %)比較有無 bias 的結果。

	Validation loss	Kaggle score (public/private)
無 bias	0.5917	0.86234/0.85743
有 bias	0.5771	0.86028/0.85371

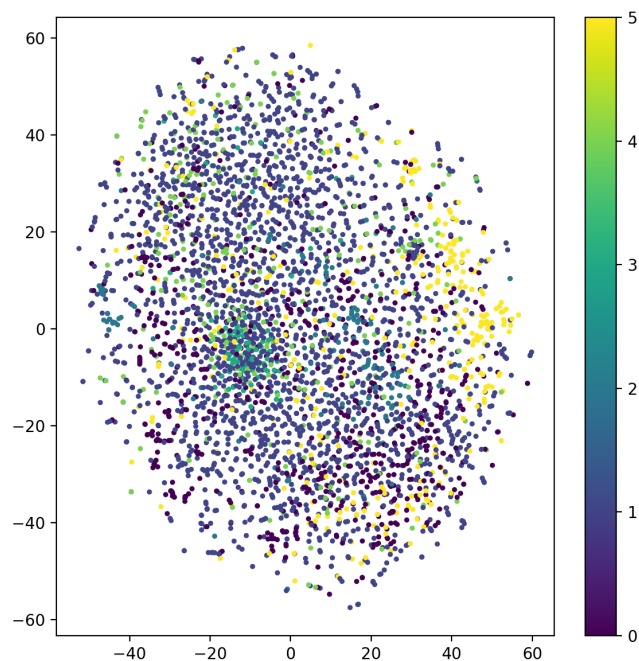
由上表可以發現有加上 bias 和沒有 bias 的 kaggle score 並沒有很大的差異，但是有加上 bias 的 model 表現稍微比沒有加上 bias 的 model 表現好。

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

我先將 movie 分成五個 category

0. Fantasy|Sci-Fi|Action|Western|Adventure|Mystery
1. Musical|Drama|Comedy|Documentary|Romance
2. Children's|Animation
3. No category
4. War|Crime|Film-Noir
5. Horror|Thriller

並用 tsne 降維後得到下圖



0 分類(紫色)偏右下角，2 分類(藍色)偏左，4 分類(淺綠色)偏上，5 分類(黃色)偏右。

5. (1%)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。

我將 movies.csv 的 ID, Genres 和 users.csv 的 ID, gender, age, occupation 加到 model 的 input，並將 model 改成 DNN 的架構(參考去年的助教課 PPT)，此方法的預測結果並沒有比我 MF model 的表現還要好，在 kaggle 上的分數只有 0.87149/0.86482(public/private)