學號：b04902053 系級：資工二 姓名：鄭淵仁

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

* normalize的方法

我總共測試兩種normalize的方法，第一種是把所有的rating直接normalize，第二種是每一個UserID都各自normalize各自的rating。

* 實驗結果

我使用相同的model，並切validation set，而且save best來測試三種不同normalize的方法。實驗結果如下表：

|  |  |  |  |
| --- | --- | --- | --- |
|  | no normalize | normalize on all rating | normalize for each user |
| validation RMSE | **0.84909(勝)** | 0.86001 | 0.86832 |

從表格中可以發現：沒有加上normalize的結果是三者中最好的。

1. **(1%)比較不同的latent dimension的結果。**

為了方便比較，我讓user和movie的embedding的latent dimension相同。

* 實驗結果

我都使用相同的model，並切validation set，而且save best來測試。

實驗結果如下圖：

從圖片中可以發現：當latent dimension降到150的時候，RMSE是最小的。

而當dimension小於150的時候，RMSE會特別大；當dimension大於150之後，RMSE會稍微增加。

* 推測原因

會有這樣的結果的原因可能是：如果latent dimension太小（<150），無法完全代表原本資料隱含的成分，結果當然很差。而當latent dimension夠大並且越來越大的時候，雖然embedding能夠代表所有資料隱含的成分了，但是因為dimension太大了，gradient descent很容易會overfit trainging set，而無法提升validation set的結果。

1. **(1%)比較有無bias的結果。**

為了方便比較，我把user和movie都各別加上bias。

* 實驗結果

我都使用相同的model，並切validation set，而且save best來測試。

實驗結果如下表：

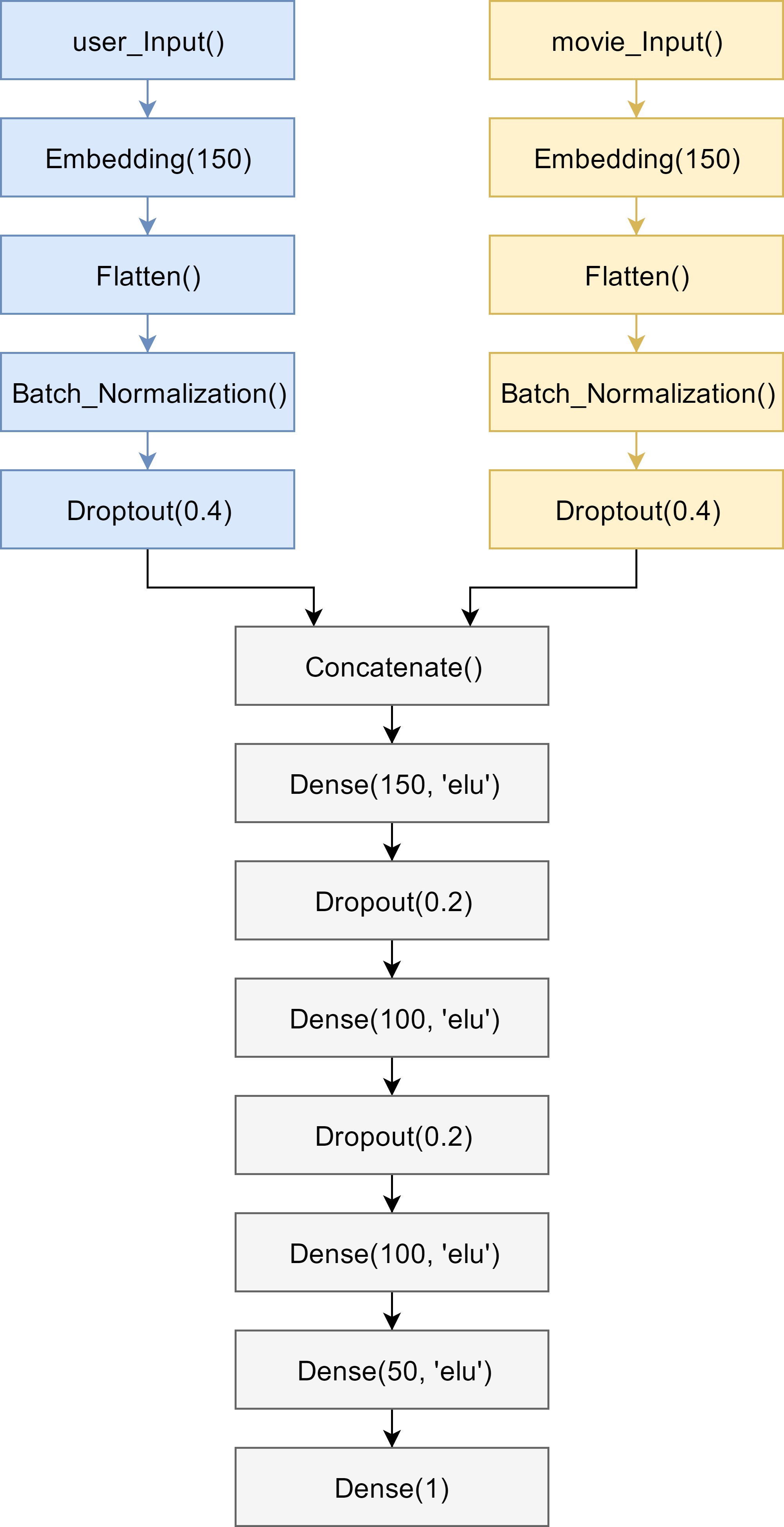
|  |  |  |
| --- | --- | --- |
|  | no bias | add bias |
| validation RMSE | **0.84909(勝)** | 0.85498 |

從表格中可以發現：沒有加上bias的結果是最好的。

1. **(1%)請試著用DNN來解決這個問題，並且說明實做的方法(方法不限)。並比較MF和NN的結果，討論結果的差異。**

* 實做的方法：

我把user的embedding和movie的embedding接在一起，再使用DNN來實作。完整的model如下圖：



* 結果及比較：

測試兩種model的時候，我都切validation set，並且save best來測試。

實驗結果如下表：

|  |  |  |
| --- | --- | --- |
|  | MF | DNN |
| validation RMSE | **0.84909(勝)** | 0.85146 |

從表格中可以發現：使用MF的結果是比較好的。

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