

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

● 標準化方法：

- 取 training set 的 mean 和 standard variance (std)，並將所有 training data 的 rating 做轉換： $\text{train} = (\text{train} - \text{mean}) / \text{std}$
- 利用轉換過的 train 進行訓練後，再將預測結果做轉換： $\text{pred} = \text{pred} * \text{std} + \text{mean}$

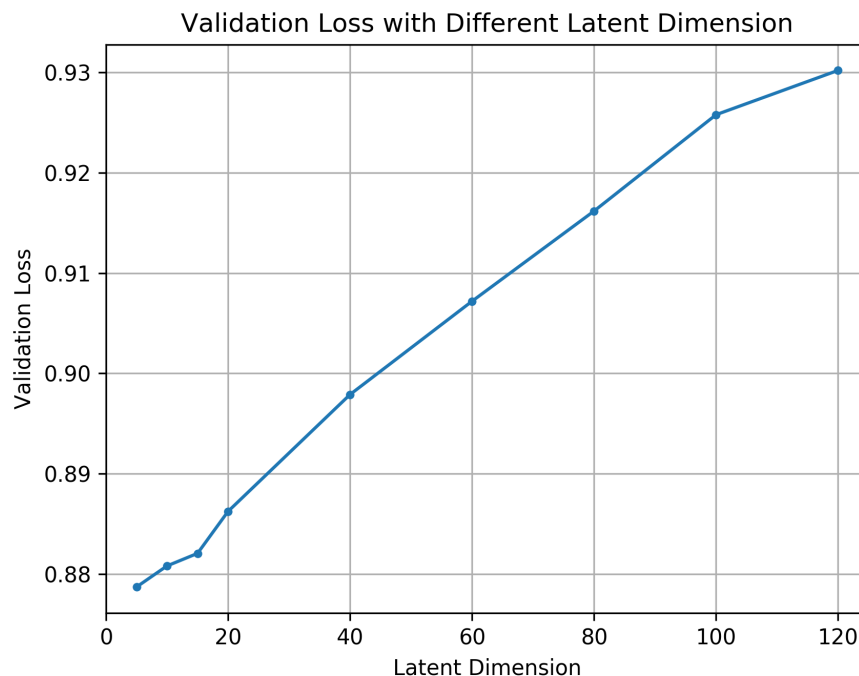
● 實驗結果：

取 MF 模型，latent dimension = 5，實驗於 testing set 的結果如下

	無標準化	有標準化
<i>Kaggle Public Score</i>	0.87531	0.86815

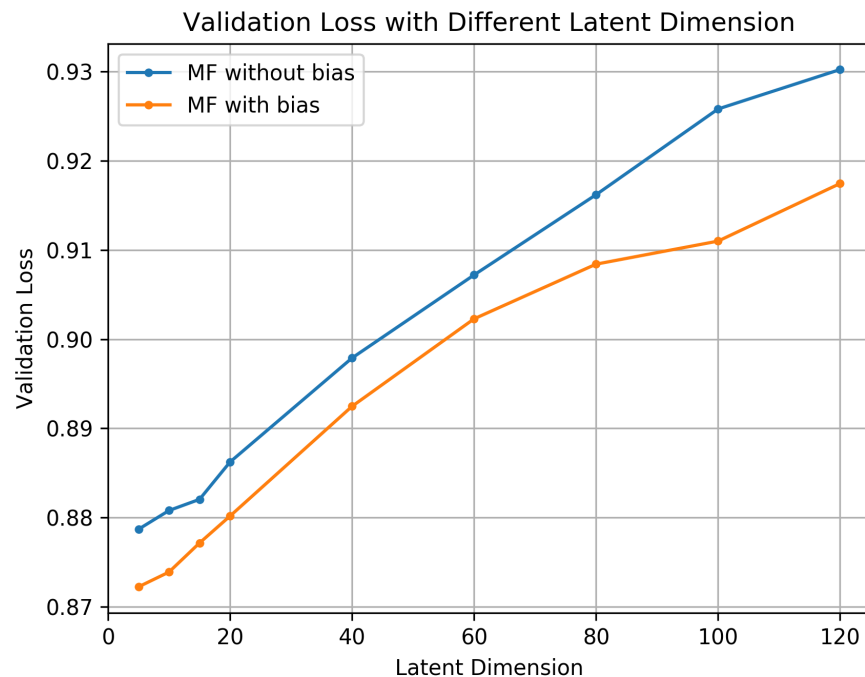
2. (1%)比較不同的 latent dimension 的結果。

未使用 normalize 及 bias，以下為不同 latent dimension 的最低 validation loss。



由上表可知，latent dimension 越小，MF 的表現越好。且實驗時發現，latent dimension 越大時，overfitting 情況越嚴重，此時 training loss 下降越快，而 validation loss 不減反增。

3. (1%)比較有無 bias 的結果。

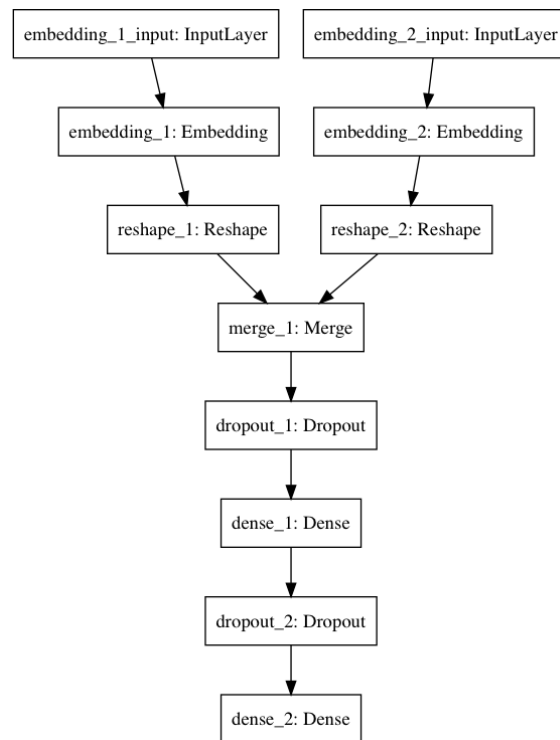


加上 bias 後 loss 降低，但仍然隨著 latent dimension 變大而變大，與沒有加 bias 的模型表現走向相同。

4. (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。

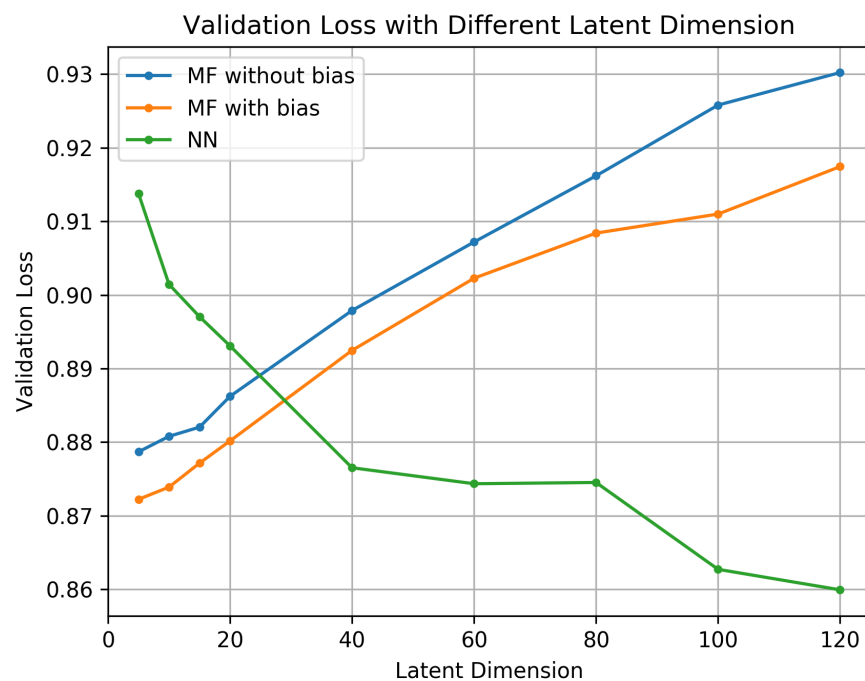
● 模型架構：

```
def dnn_model(n_users, m_items, k_factors):  
    P = Sequential()  
    P.add(Embedding(n_users, k_factors, input_length = 1))  
    P.add(Reshape((k_factors,)))  
    Q = Sequential()  
    Q.add(Embedding(m_items, k_factors, input_length = 1))  
    Q.add(Reshape((k_factors,)))  
  
    model = Sequential()  
    model.add(Merge([P, Q], mode = 'concat'))  
    model.add(Dropout(0.3))  
    model.add(Dense(k_factors, activation = 'relu'))  
    model.add(Dropout(0.3))  
    model.add(Dense(1, activation = 'linear'))  
  
    model.compile(loss = 'mse', optimizer = 'nadam', metrics = [])  
    model.summary()  
  
    return model
```



- 差異比較：

下表比較不同 latent dimension 對此模型的影響



觀察上表可發現，使用較小 latent dimension 時，MF 有較好的表現，NN 則需要較大的 latent dimension 來降低 loss。

5. (1%)請試著將 movie 的 embedding 用 tsne 降維後，將 movie category 當作 label 來作圖。

我的電影分類如下：

```
m_class.append(['Animation', "Children's", 'Comedy'])
m_class.append(['Thriller', 'Horror'])
m_class.append(['Western', 'Action', 'Crime', 'War'])
m_class.append(['Adventure', 'Fantasy', 'Sci-Fi'])
m_class.append(['Drama', 'Documentary'])
m_class.append(['Romance', 'Musical'])
m_class.append(['Film-Noir', 'Mystery'])
```

作圖結果如下：

