深度学习课程作业

结题报告

Background—The Netflix Prize

From 2006 to 2009, Netflix sponsored a competition, offering a grand prize of $1,000,000 to the team that could take an offered dataset of over 100 million movie ratings and return recommendations that were 10% more accurate than those offered by the company's existing recommender system.

Dataset

	Movie1	Movie2	Movie3	Movie4	...	Movie17,770
User1	5	?	?	1	...	?
User2	3	?	2	?	...	?
User3	?	?	?	4	...	2
User4	?	3	?	?	...	?
...	...	...	...	...	...	...
User48,189	3	?	?	?	...	?

总共
100,480,507
个评分

矩阵元素总
共:
48,189*17,7
70=
856,318,530

矩阵密度:
0.117

Sampling

- Netflix给出的数据集共有17,770部电影；48,189个用户；100,480,507个评分

| | movieRating.db | 2021/10/25 20:19 | DB 文件 | 2,757,824 KB |

- 在其中选取了10,000个用户的数据 1,048,574个评分

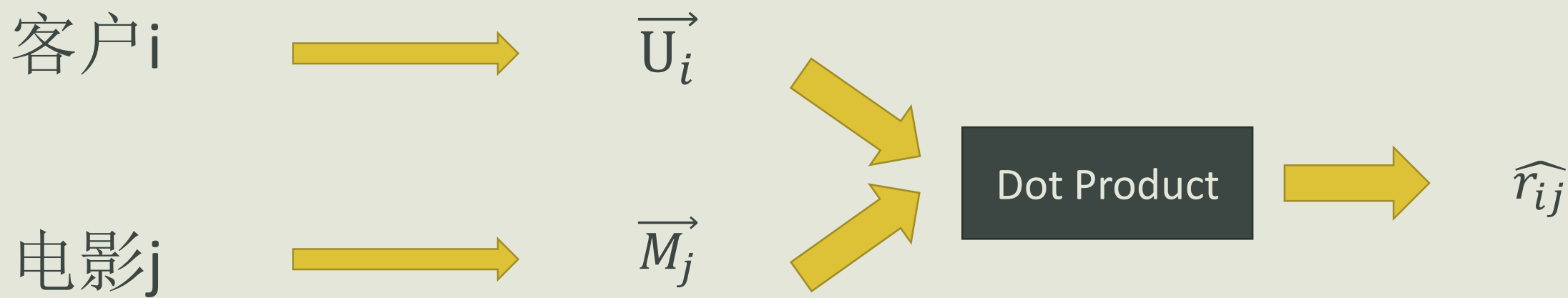| D	E	F	G	H	
movieId	rating	timestamp	userId_nor	movieId_nor	
30	3	#######	0	0	
157	3	#######	0	1	
173	4	#######	0	2	
175	5	#######	0	3	
191	2	#######	0	4	
197	3	#######	0	5	
241	3	#######	0	6	
295	4	#######	0	7	
299	3	#######	0	8	
329	4	#######	0	9	

Training set: 90%
Validation set: 10%

Baseline&Goal

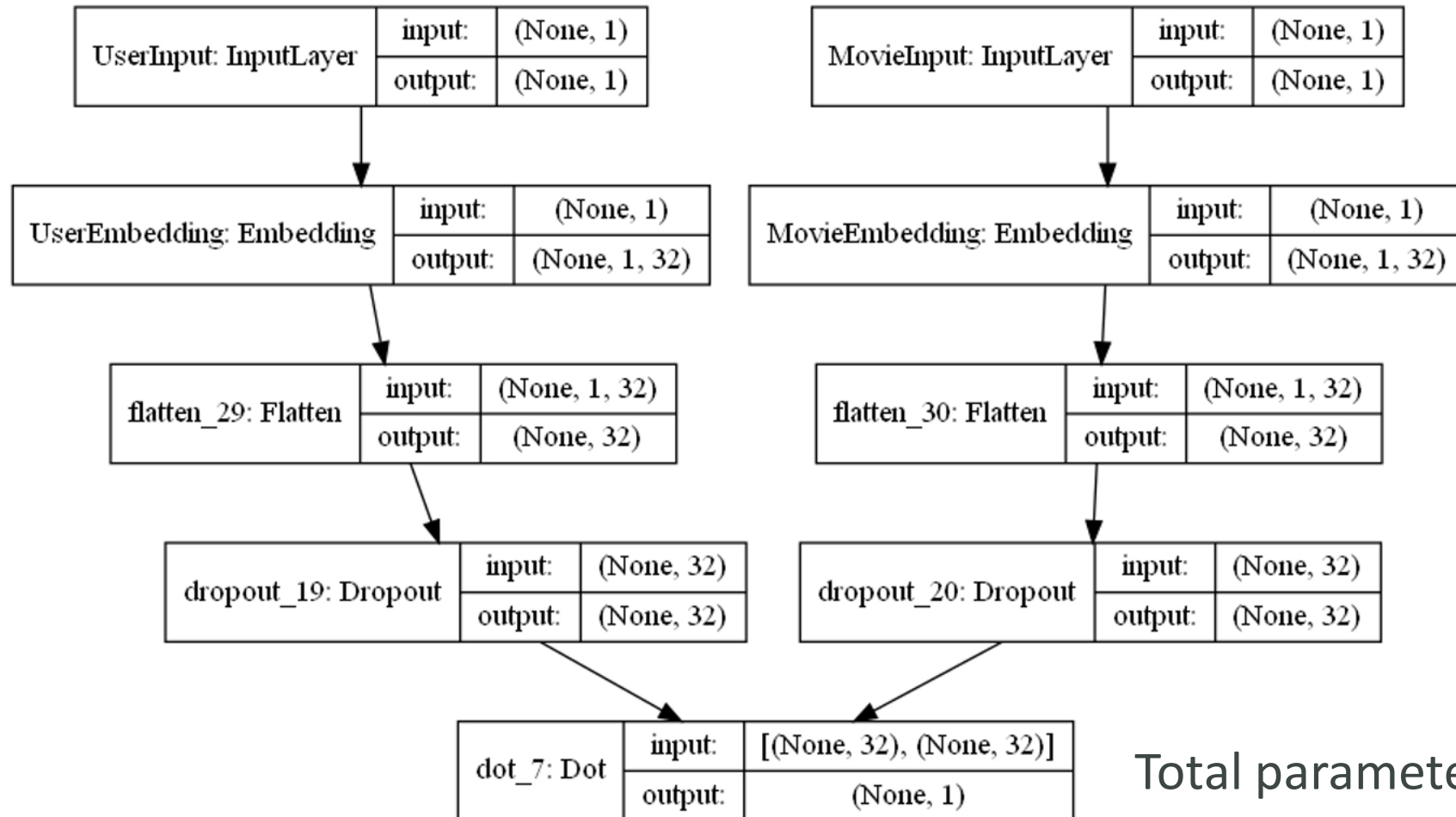$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}.$$

- 用RMSE作为loss function

Netflix	10% Improvement	2007 Progress Prize	2008 Progress Prize	2009 Grand Prize
0.9514	0.8563	0.8712	0.8616	0.8554

MF model

客户i $\longrightarrow$ $\overrightarrow{U_i}$

电影j $\longrightarrow$ $\overrightarrow{M_j}$

Dot Product $\longrightarrow$ $\widehat{r_{ij}}$

MF model implemented with Keras

Total parameter:679,424

The performance of MF

	Train RMSE	Valid RMSE
MF	2.915	2.923
MF with normalization	0.912	0.931
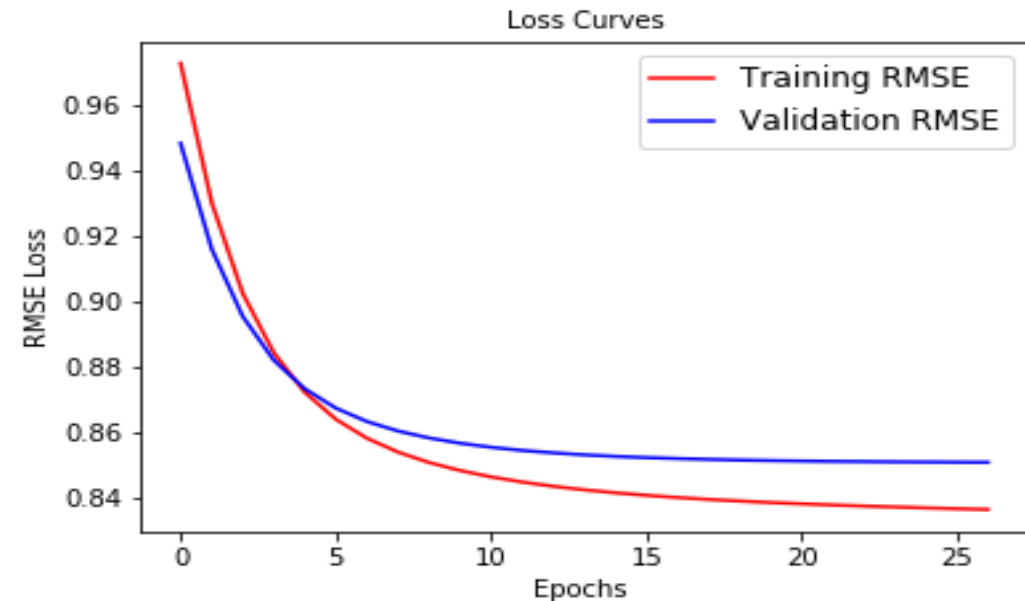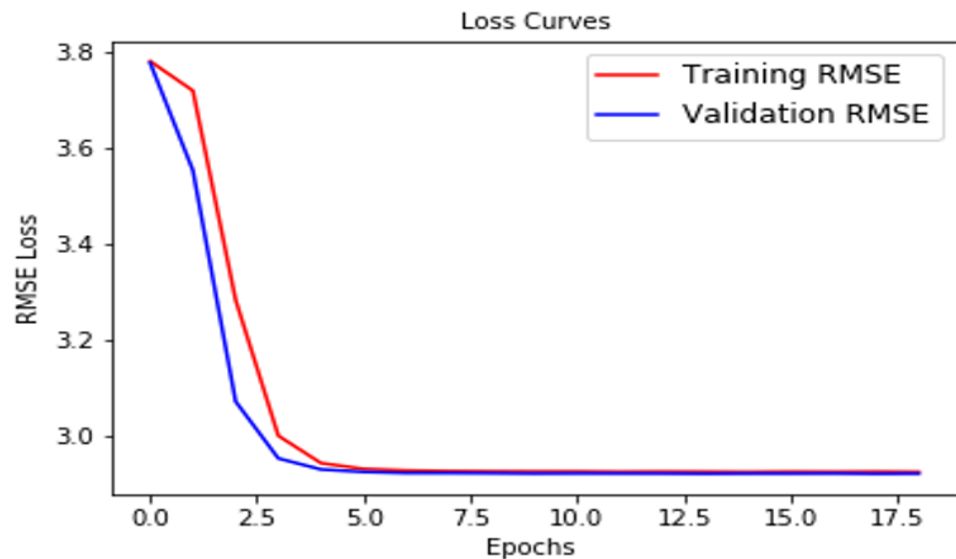
The performance of MF

	Train RMSE	Valid RMSE
MF	2.915	2.923
MF with normalization	0.912	0.931

Netflix	10% Improvement	2007 Progress Prize	2008 Progress Prize	2009 Grand Prize
0.9514	0.8563	0.8712	0.8616	0.8554

MF model

客户i $\longrightarrow$ $\overrightarrow{U_i}$

电影j $\longrightarrow$ $\overrightarrow{M_j}$

Dot Product $\longrightarrow$ $\widehat{r_{ij}}$

DL model

客户i $\longrightarrow$ $\overrightarrow{U_i}$

电影j $\longrightarrow$ $\overrightarrow{M_j}$ $\longrightarrow$ NN $\longrightarrow$ $\widehat{r_{ij}}$
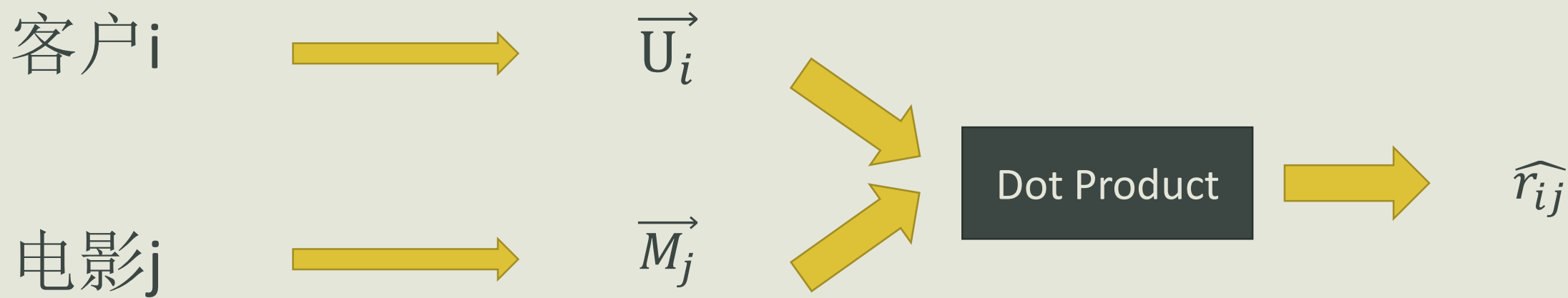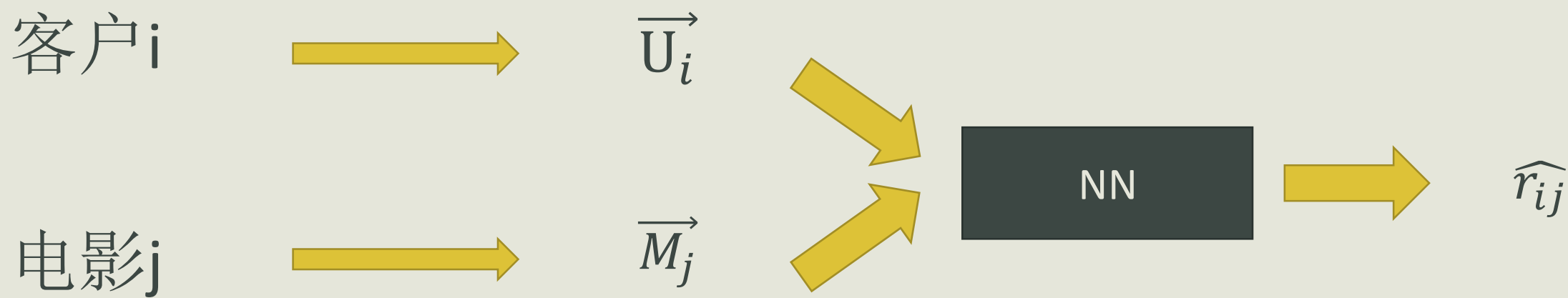
Deep Learning Model

Total Parameter
782,529
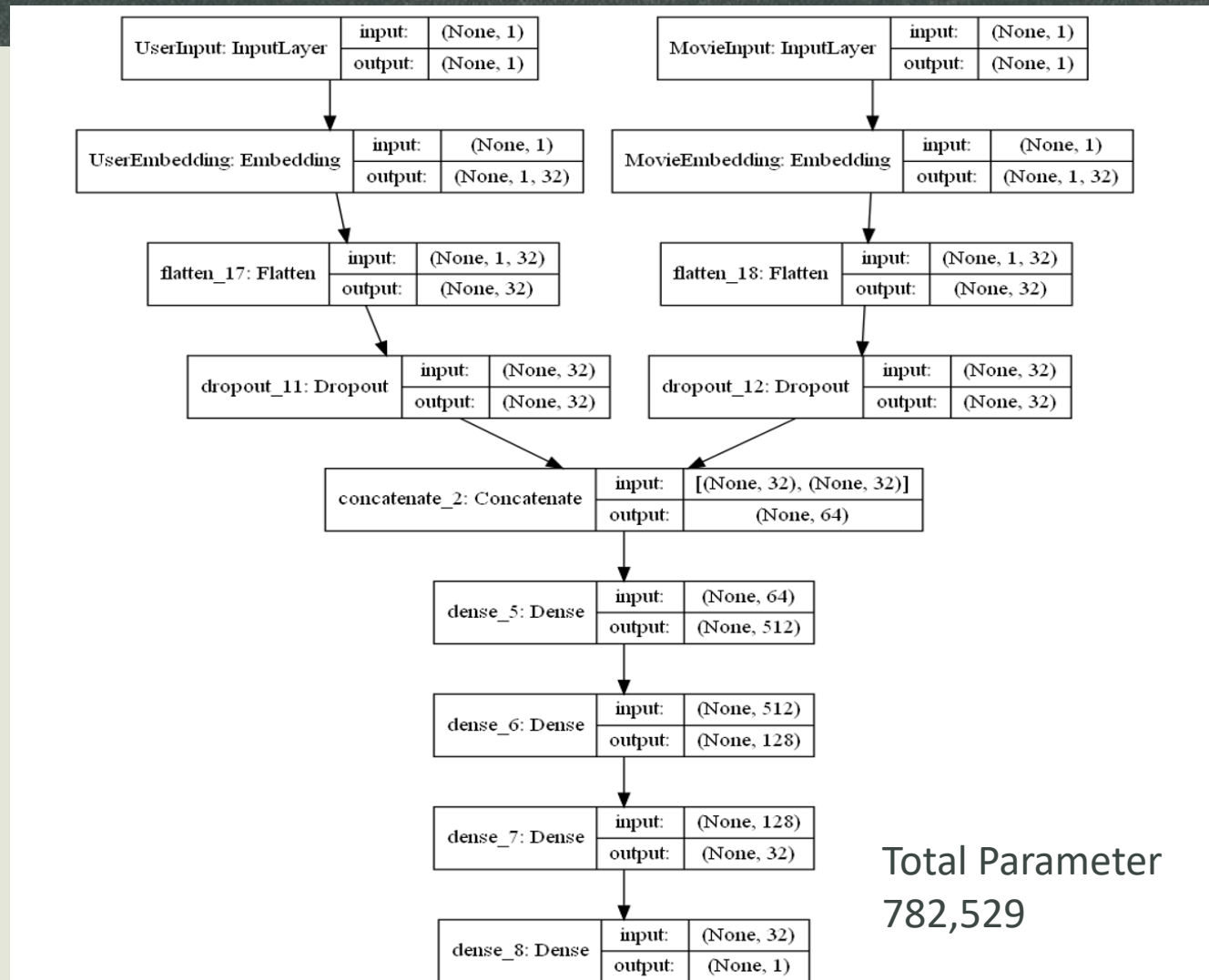
Performance of DL model

	Train RMSE	Valid RMSE
DL	0.8235	0.878
DL with Normalization	0.831	0.882
MF	2.915	2.923
MF with Normalization	0.912	0.931
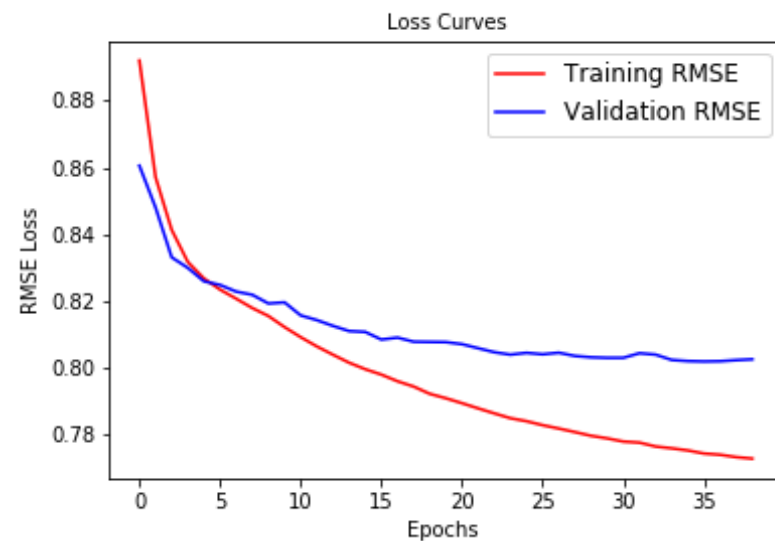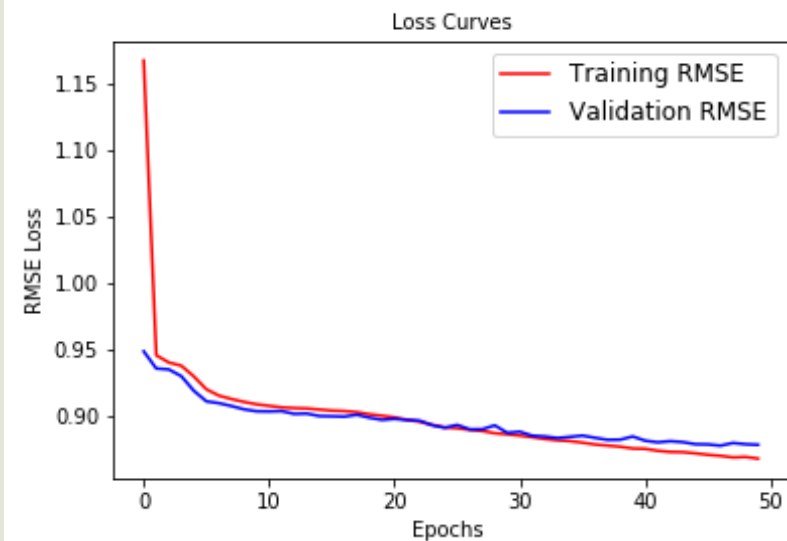
DL model– Adding Bias
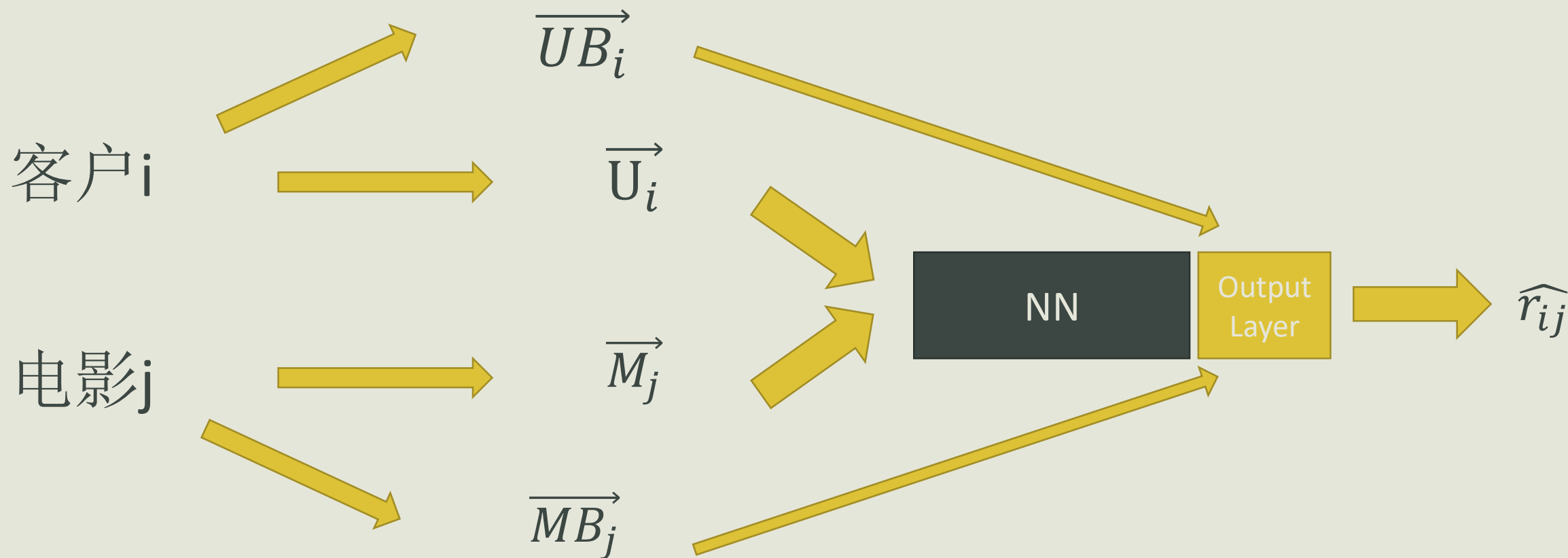
Performance of DL model

	Train RMSE	Valid RMSE
DL	0.824	0.878
DL with Normalization	0.831	0.882
DL with Bias	0.817	0.873
DL with Normalization&Bias	0.806	0.871
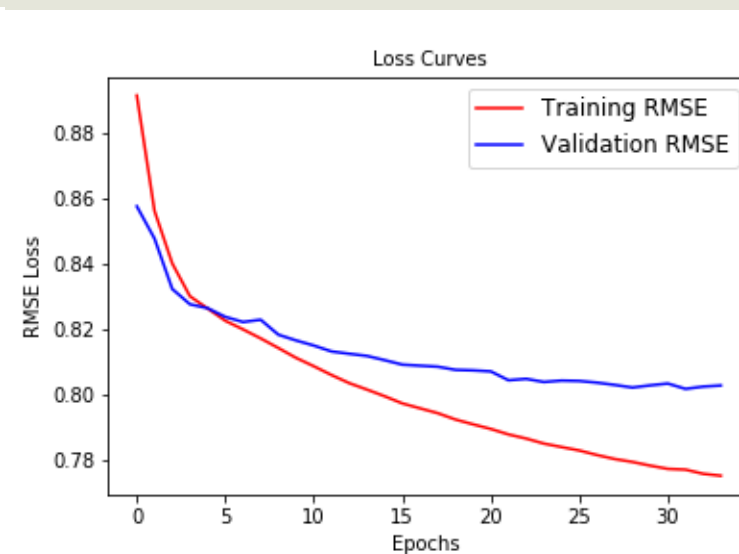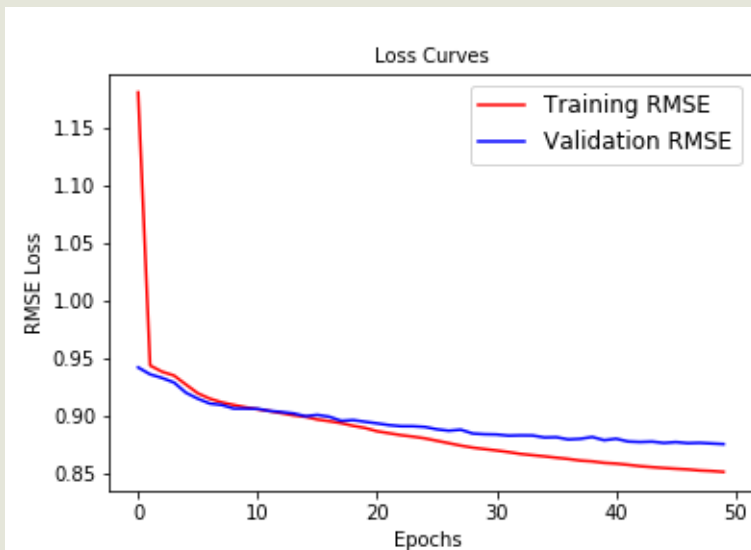
Performance of DL model

Final Result

Train RMSE	Valid RMSE
0.806	0.871

Train time: 10s per epoch

Netflix	10% Improvement	2007 Progress Prize	2008 Progress Prize	2009 Grand Prize
0.9514	0.8563	0.8712	0.8616	0.8554

Thanks