Dive into Deep Learning

Chapter 16. Deep Factorization Machines

Wayne L, Mar, 14

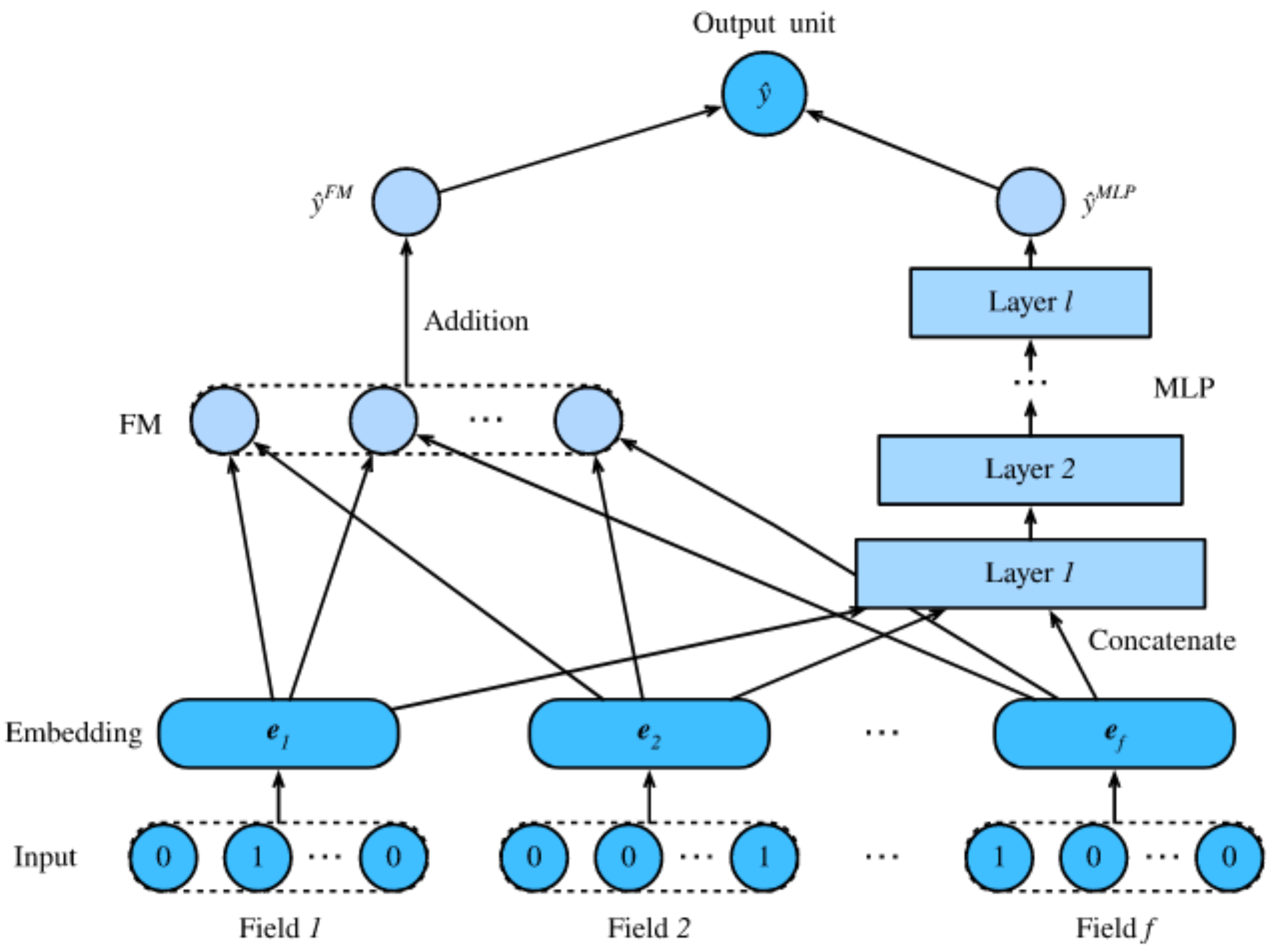
16.10 Deep Factorization Machines

- Learning effective feature combinations is critical to the success of CTR prediction task.
- Factorization machines model feature interactions in a linear paradigm (e.g., bilinear paradigm)
- This is often insufficient for real-word data where inherent feature crossing structures are usually very complex and nonlinear. What's worse, second-order feature interactions are generally used in factorization machines in practice. Modeling higher degrees of feature combinations with factorization machines is possible theoretically but it is usually not adopted due to numerical instability and high computational complexity.
- Deep neural networks are powerful in feature representation learning and have the potential to learn sophisticated feature interactions. As such, it is natural to integrate deep neural networks to factorization machines. Adding nonlinear transformation layers to factorization machines gives it the capability to model both low-order feature combinations and high-order feature combinations. Moreover, non-linear inherent structures from inputs can also be captured with deep neural networks.
- DeepFM (https://d2l.ai/chapter_references/zreferences.html#guo-tang-ye-ea-2017

16.10 Deep Factorization Machines

Model Architecture

- DeepFM consists of an FM component and a deep component which are integrated in a parallel structure.
- The FM component is the same as the 2-way factorization machines which is used to model the low-order feature interactions.
- The deep component is a MLP that is used to capture high-order feature interactions and nonlinearities.
- These two components share the same inputs/embeddings and their outputs are summed up as the final prediction (CTR).



16.10 Deep Factorization Machines

Model Architecture

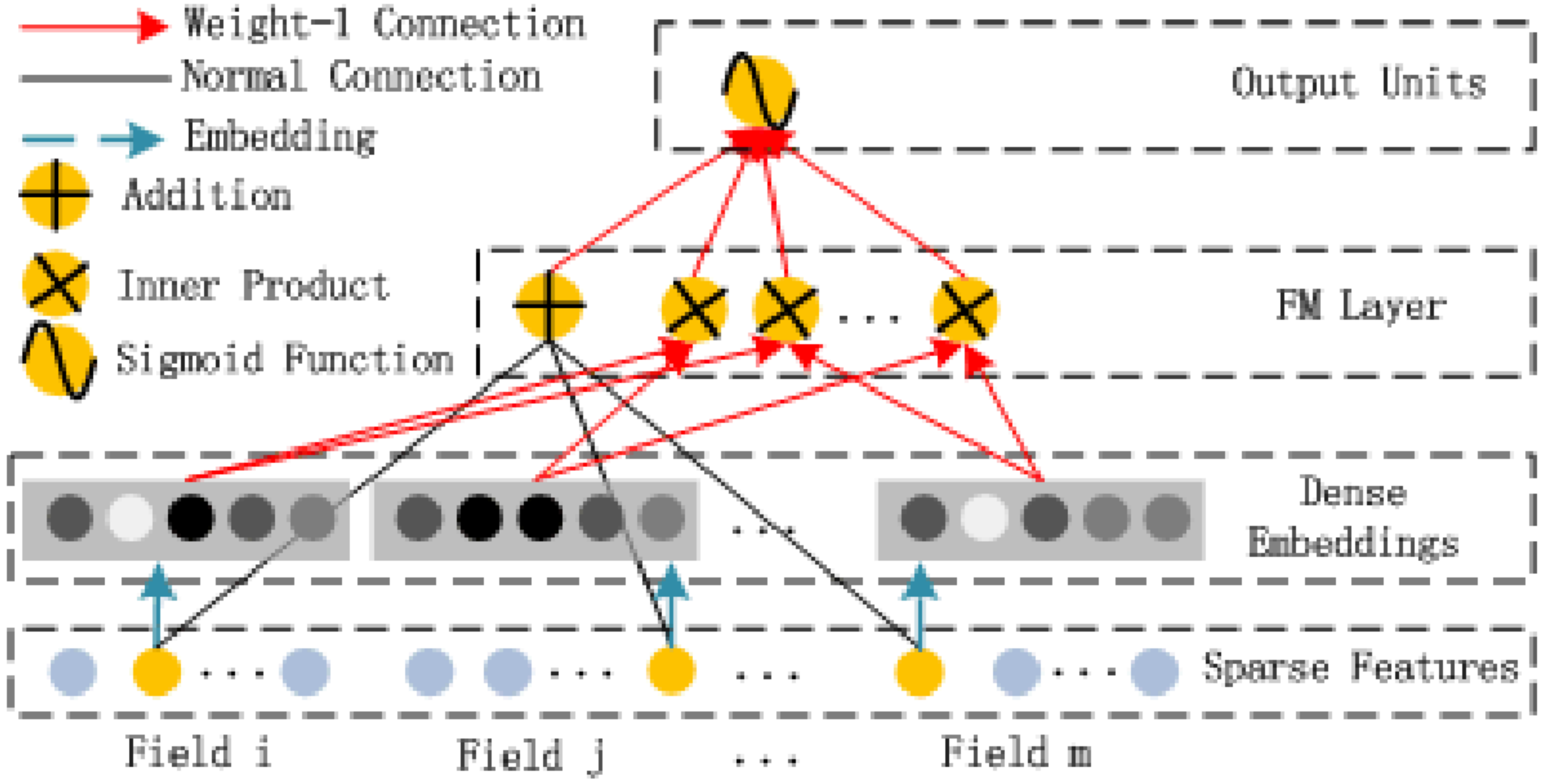
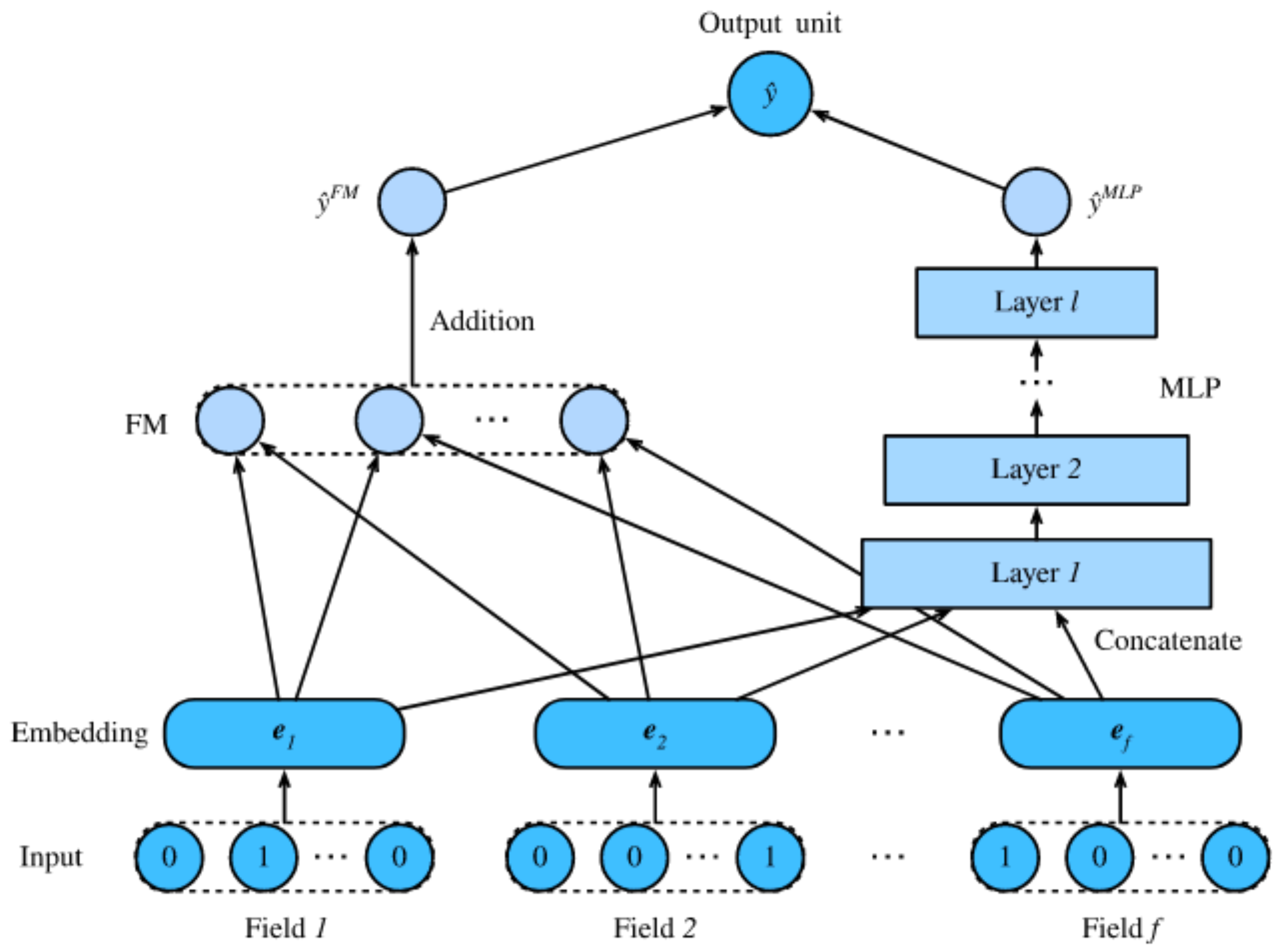


Figure 2: The architecture of FM.



16.10 Deep Factorization Machines

Dataset

Table 2: Performance on CTR prediction.

	Company∗		Criteo	
	AUC	LogLoss	AUC	LogLoss
LR	0.8640	0.02648	0.7686	0.47762
FM	0.8678	0.02633	0.7892	0.46077
FNN	0.8683	0.02629	0.7963	0.45738
IPNN	0.8664	0.02637	0.7972	0.45323
OPNN	0.8658	0.02641	0.7982	0.45256
PNN∗	0.8672	0.02636	0.7987	0.45214
LR & DNN	0.8673	0.02634	0.7981	0.46772
FM & DNN	0.8661	0.02640	0.7850	0.45382
DeepFM	**0.8715**	**0.02618**	**0.8007**	**0.45083**

- Criteo dataset
  - 45M user's click records
  - 13 continuous features + 26 categorical features
  - Train: 90%, Test: 10%

- Company dataset (huawei)
  - Company App Store
  - 7 consecutive days of user's click records for training, next 1 day for testing
  - 1 billion records
  - app features (category etc) + user features (user's downloaded apps) + context features (operation time)