

Outbrain Click Prediction

Team: 1丁讚讚讚

Outline

- Model
- Feature Extraction
- Evaluation
- Experiments

Model

- Factorization Machine (FM)
- Weighted FM
- List-wise Framework

Factorization Machine

objective function

$$\min_{\mathbf{w}} - \sum_{i=1}^L (y_i \log p + (1-y_i) \log (1-p)) + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

$$p = \text{sigmoid}(\phi(\mathbf{w}, \mathbf{x}))$$

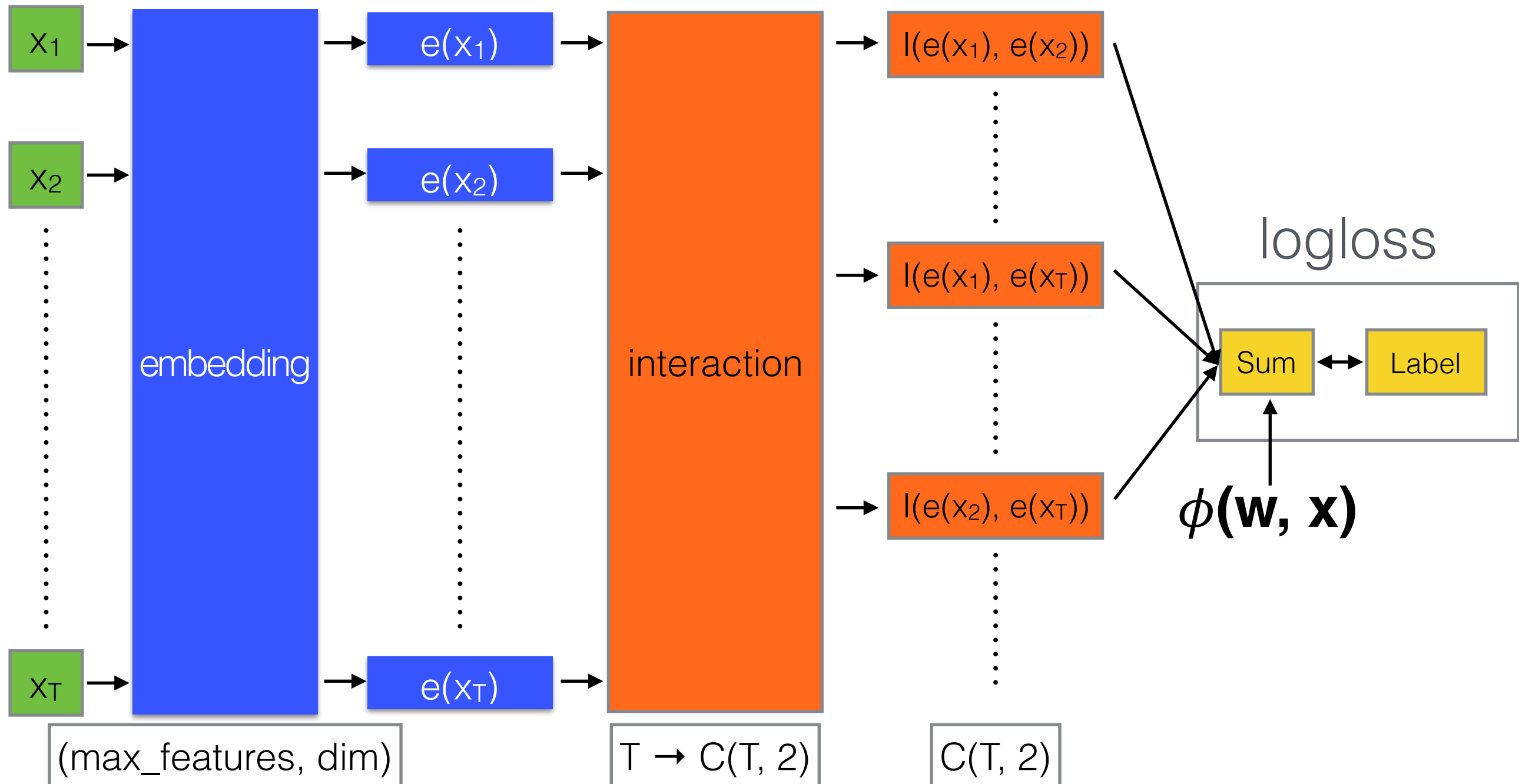
$$\phi(\mathbf{w}, \mathbf{x}) = \sum_{j_1, j_2 \in C_2} \langle \mathbf{w}_{x_{j_1}}, \mathbf{w}_{x_{j_2}} \rangle$$

L: number of instances

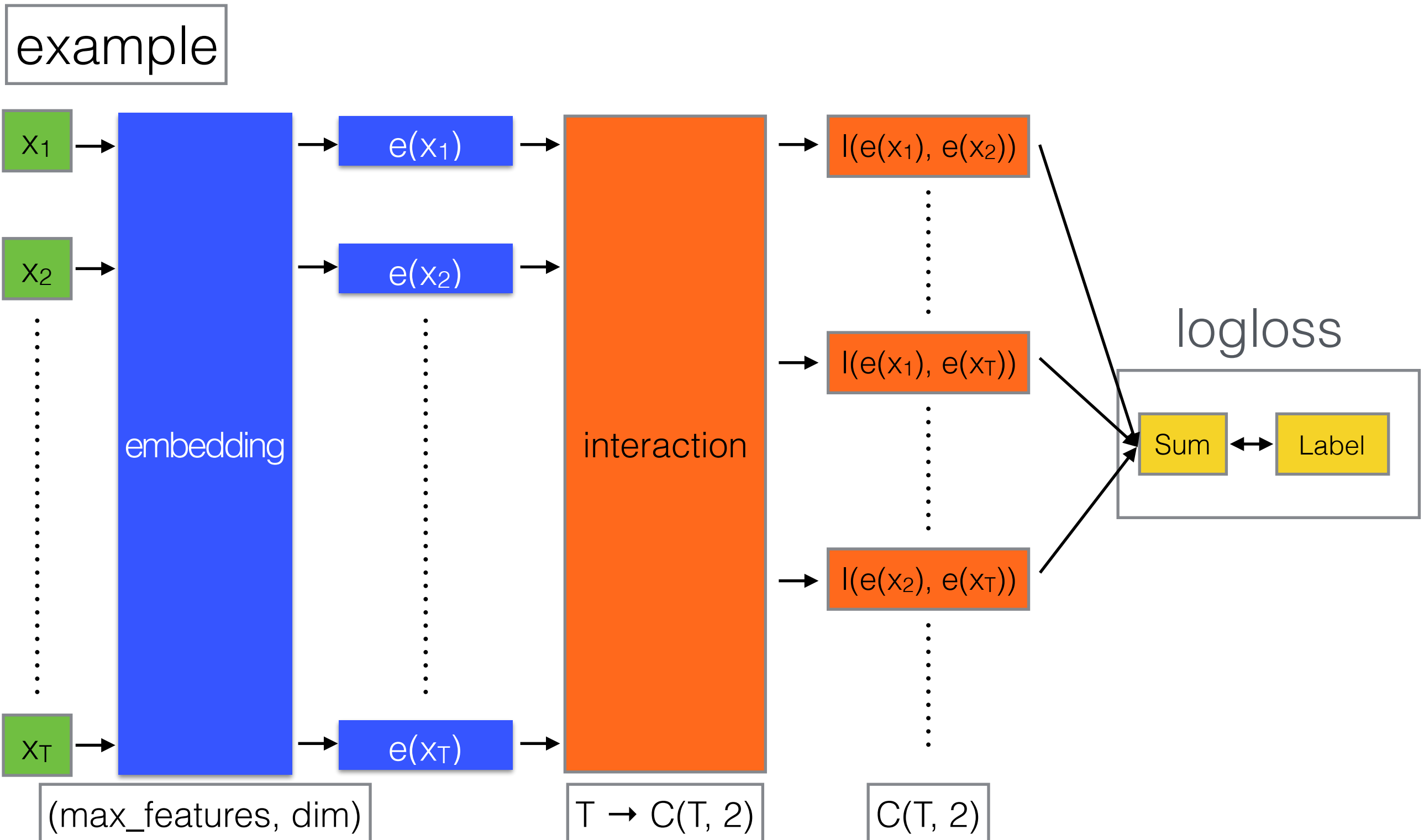
λ : regularization parameter

Factorization Machine

illustration



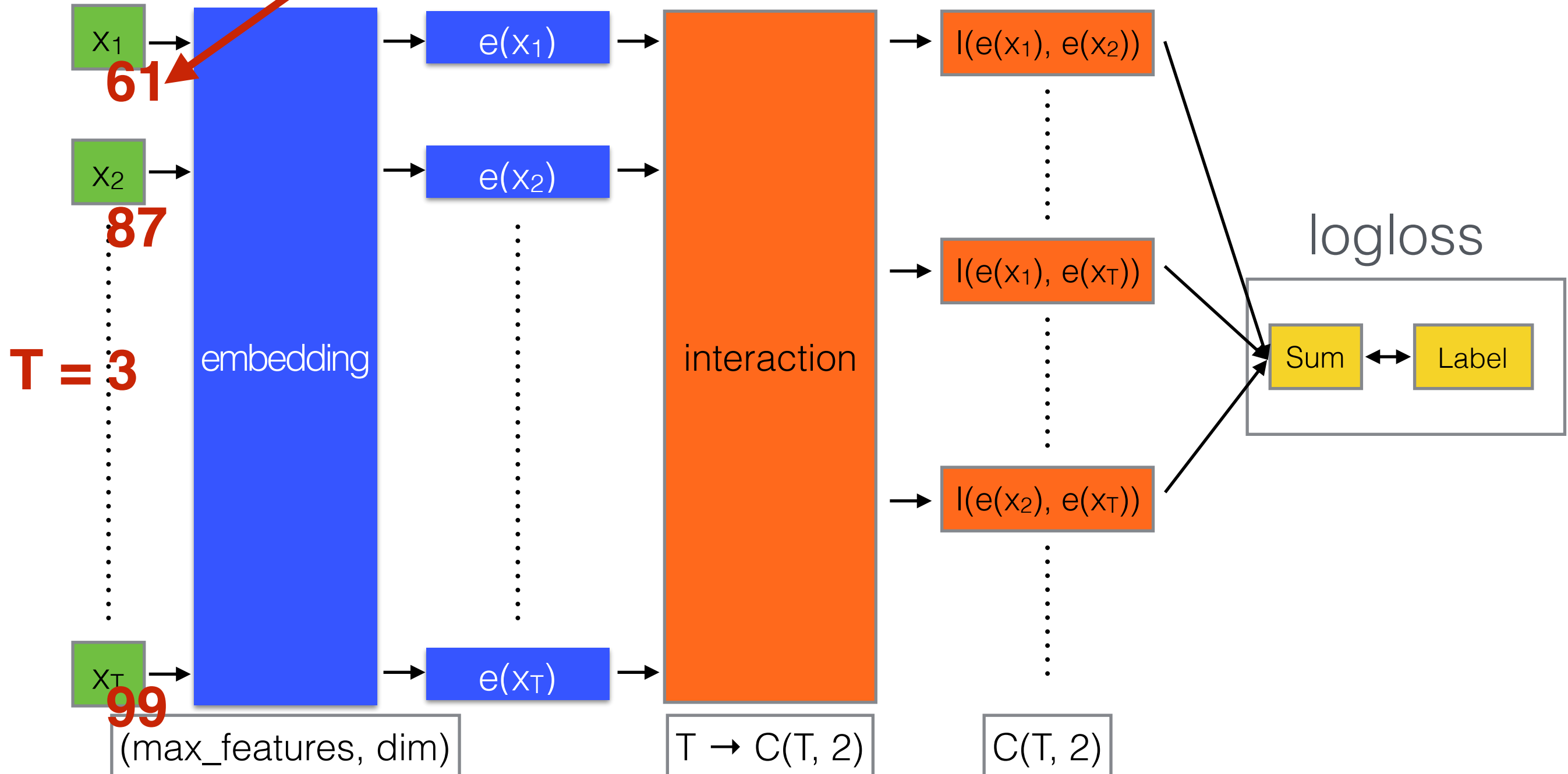
Factorization Machine



Factorization Machine

example

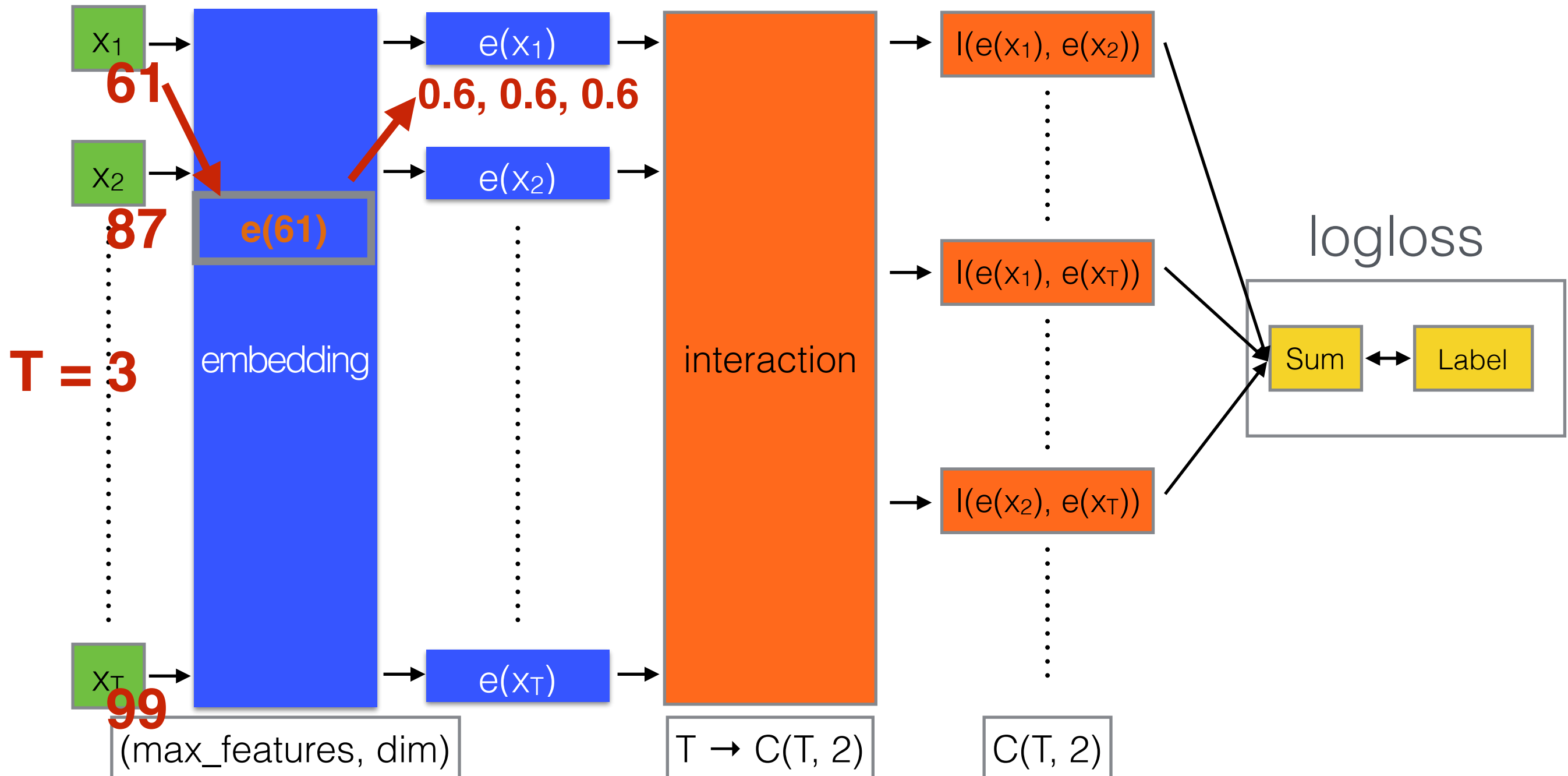
feature 61 = 1



Factorization Machine

example

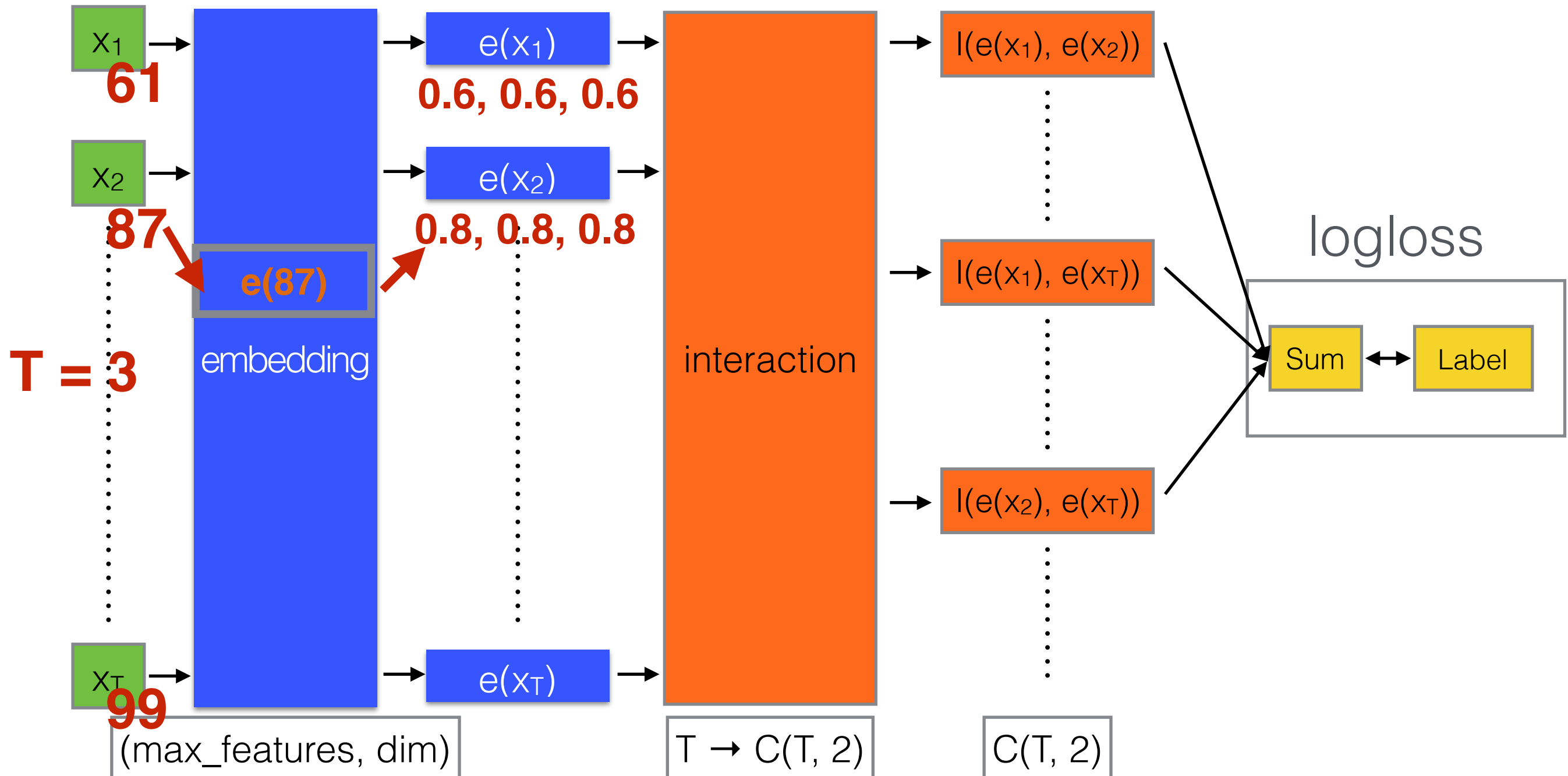
dim = 3



Factorization Machine

example

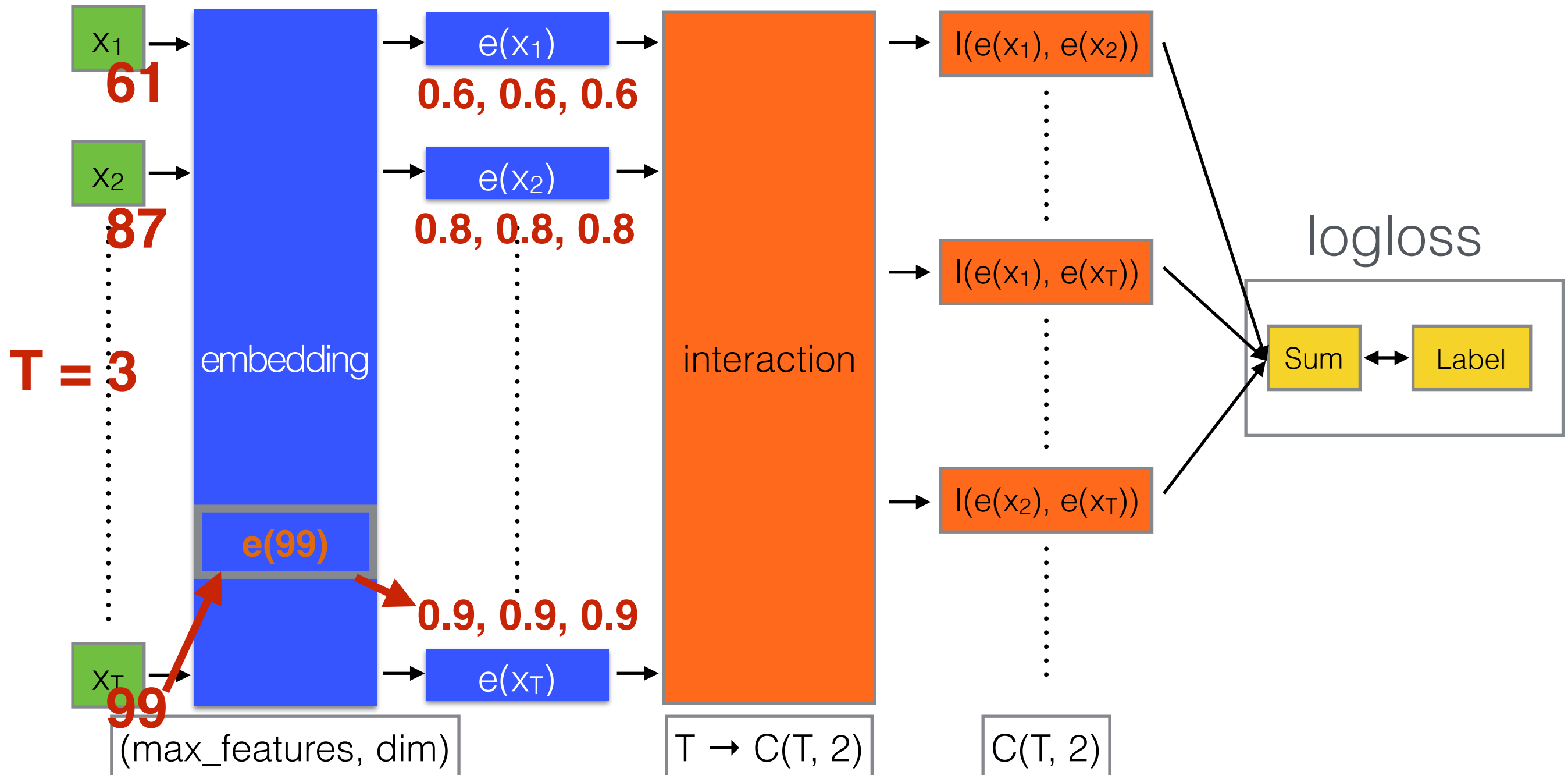
dim = 3



Factorization Machine

example

dim = 3

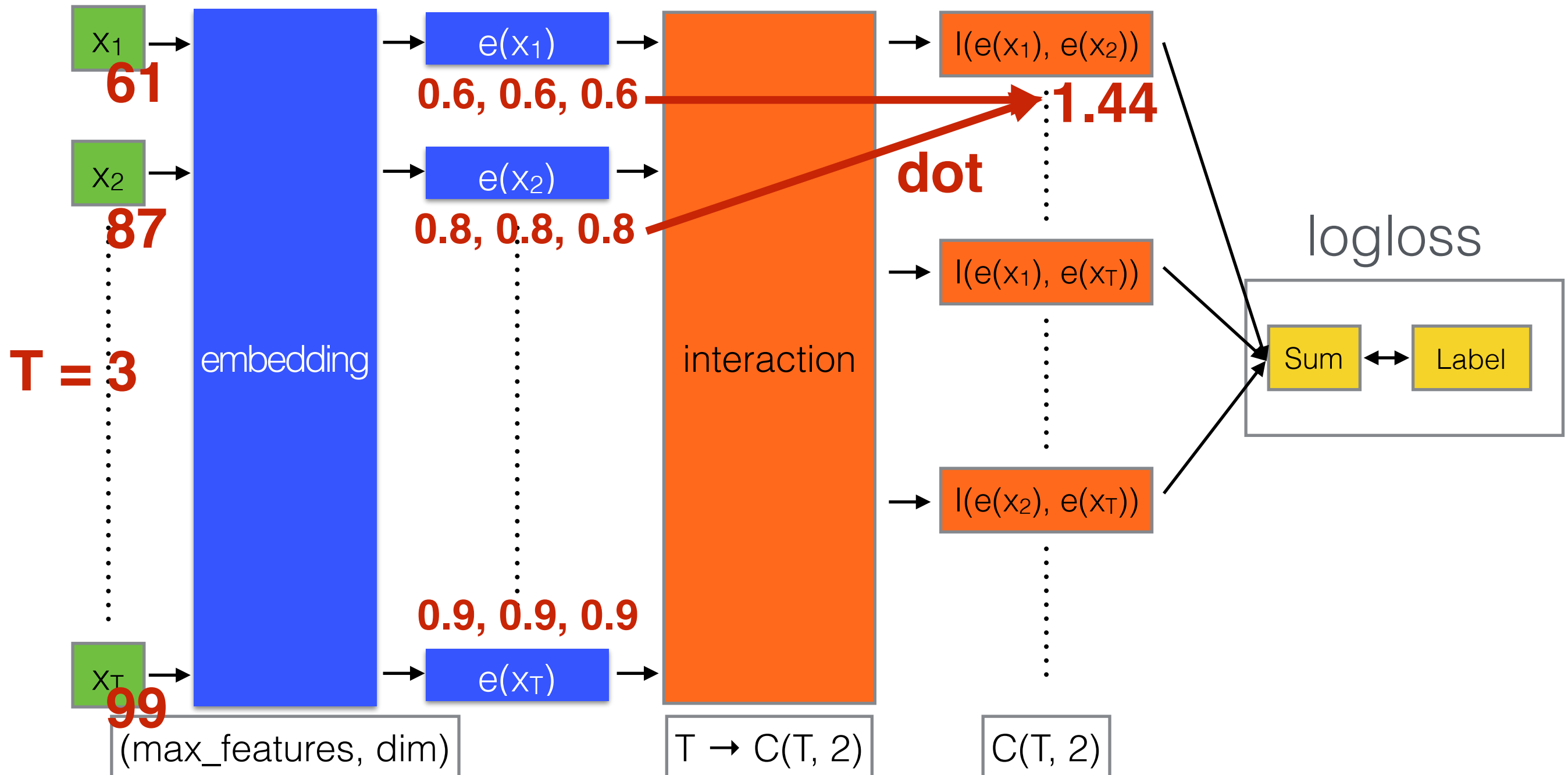


Factorization Machine

example

dim = 3

$C(3, 2) = 3$

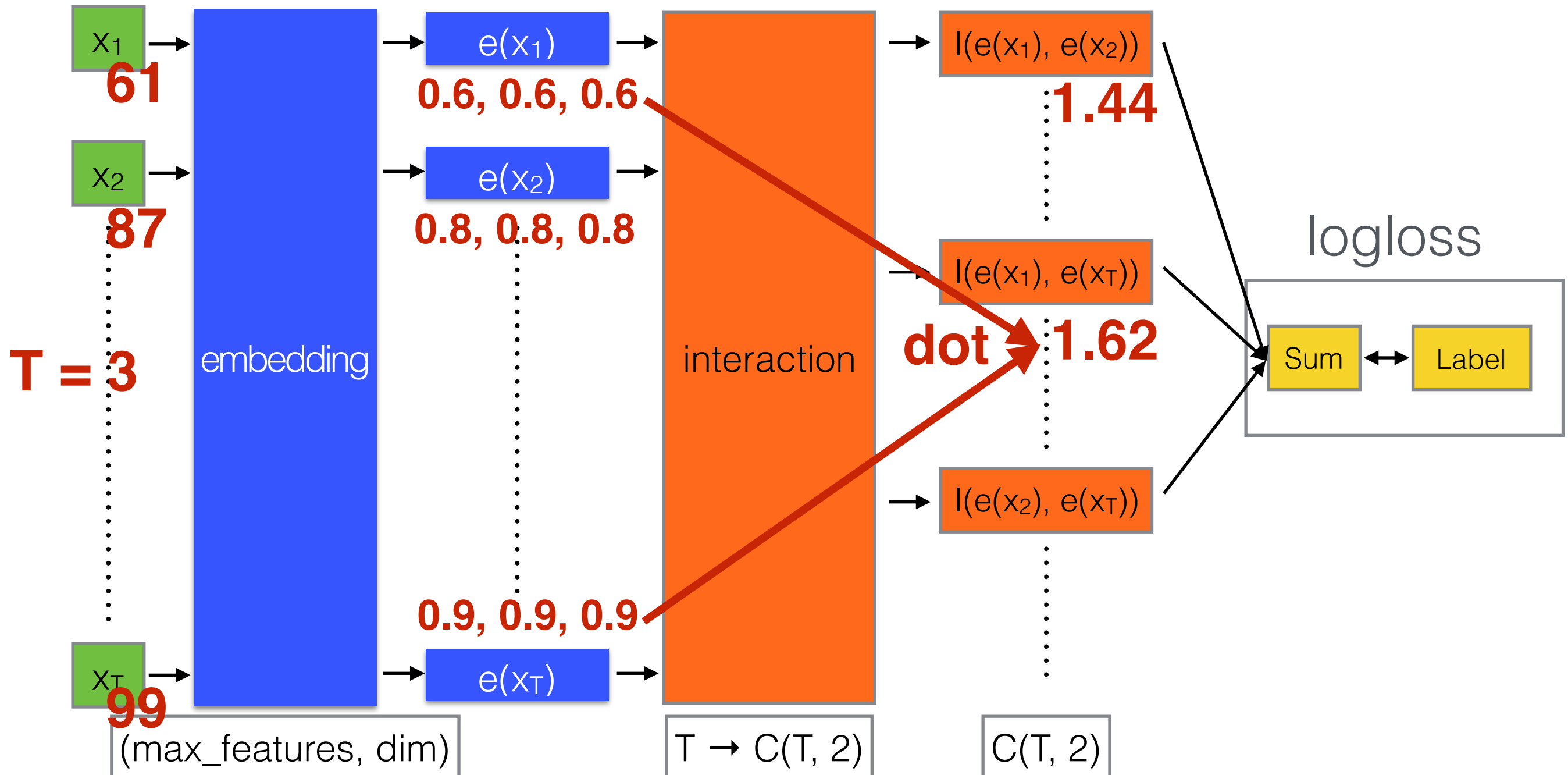


Factorization Machine

example

dim = 3

$C(3, 2) = 3$

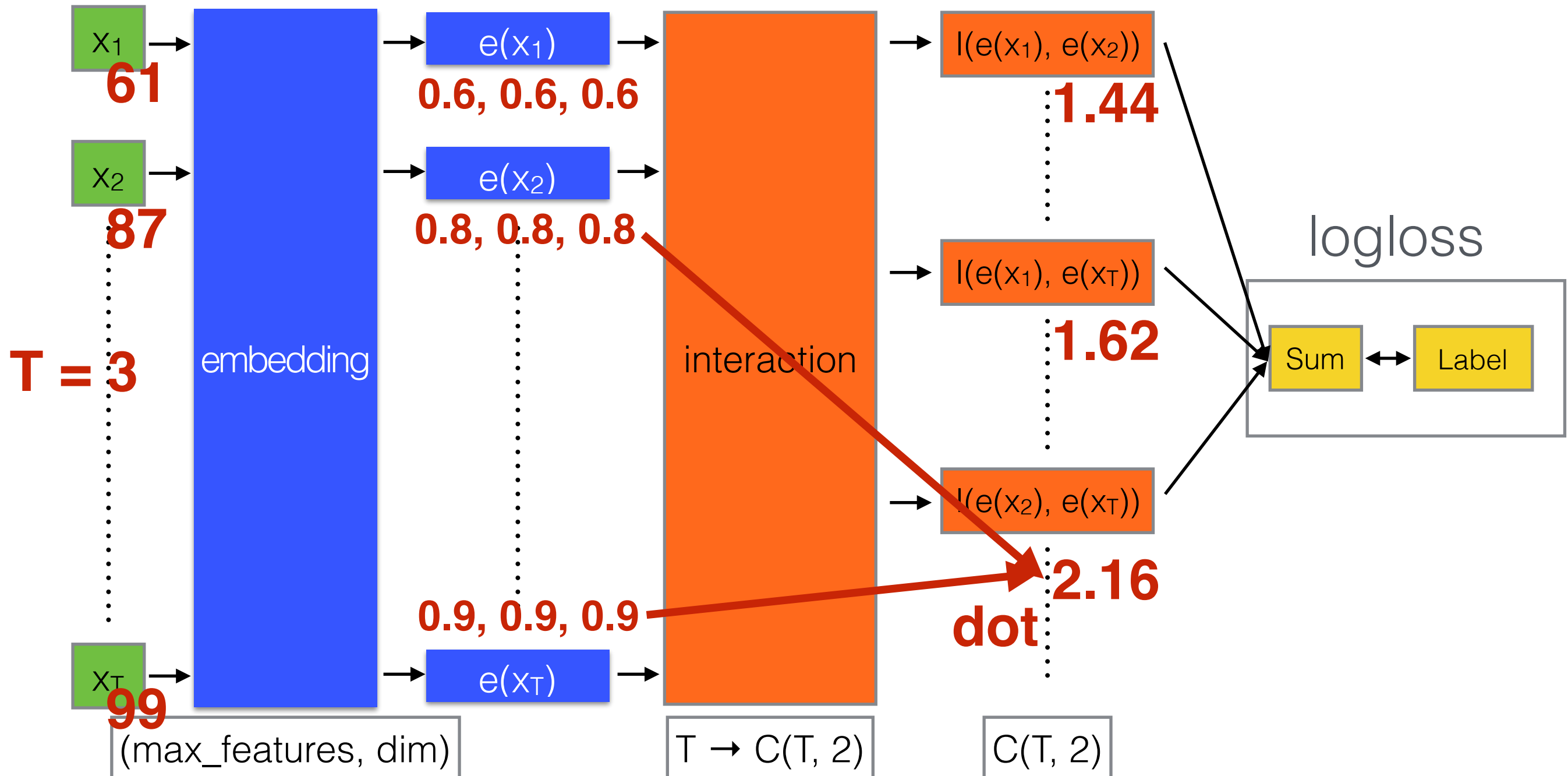


Factorization Machine

example

dim = 3

$C(3, 2) = 3$

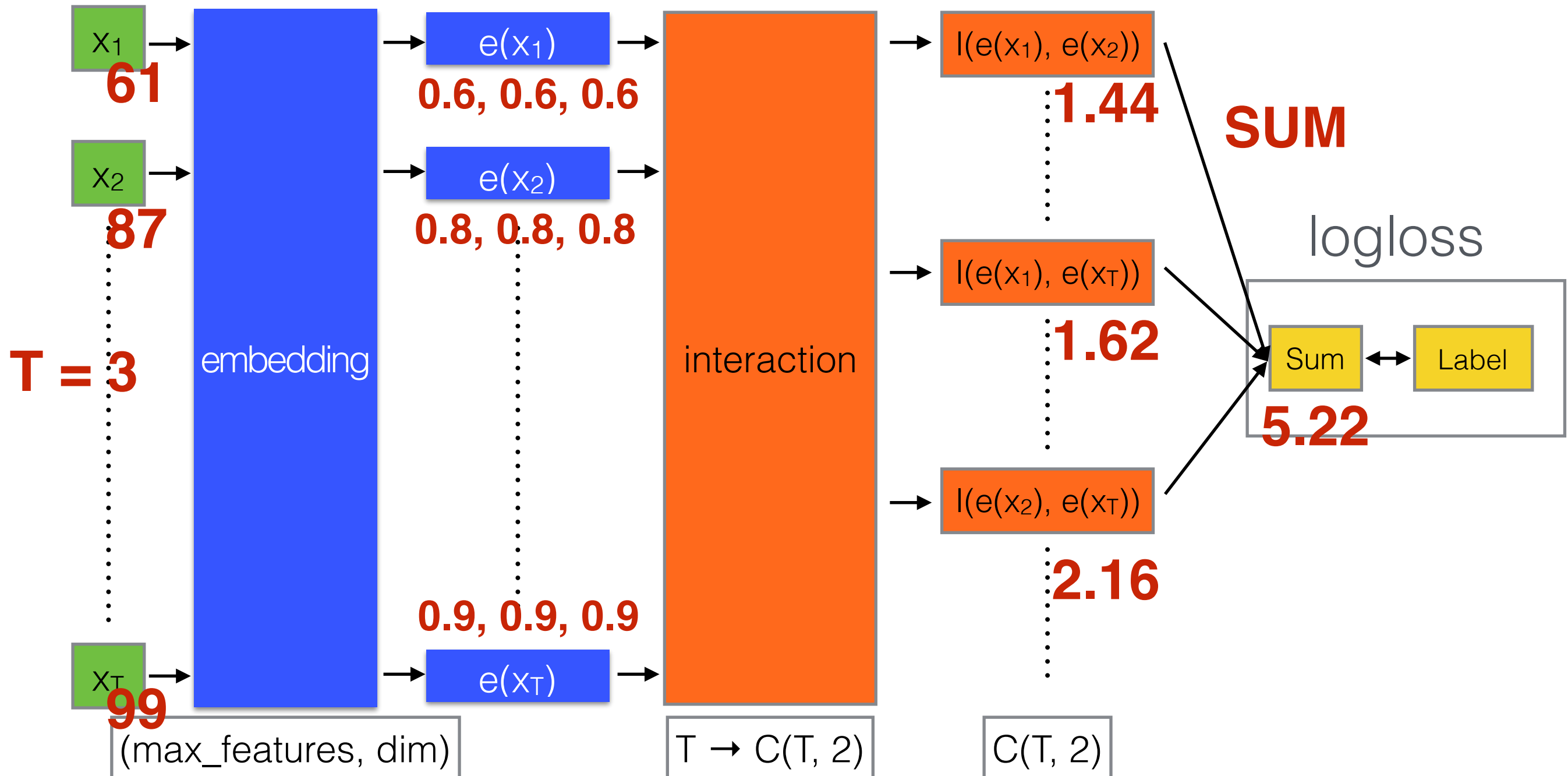


Factorization Machine

example

dim = 3

$C(3, 2) = 3$



Weighted FM

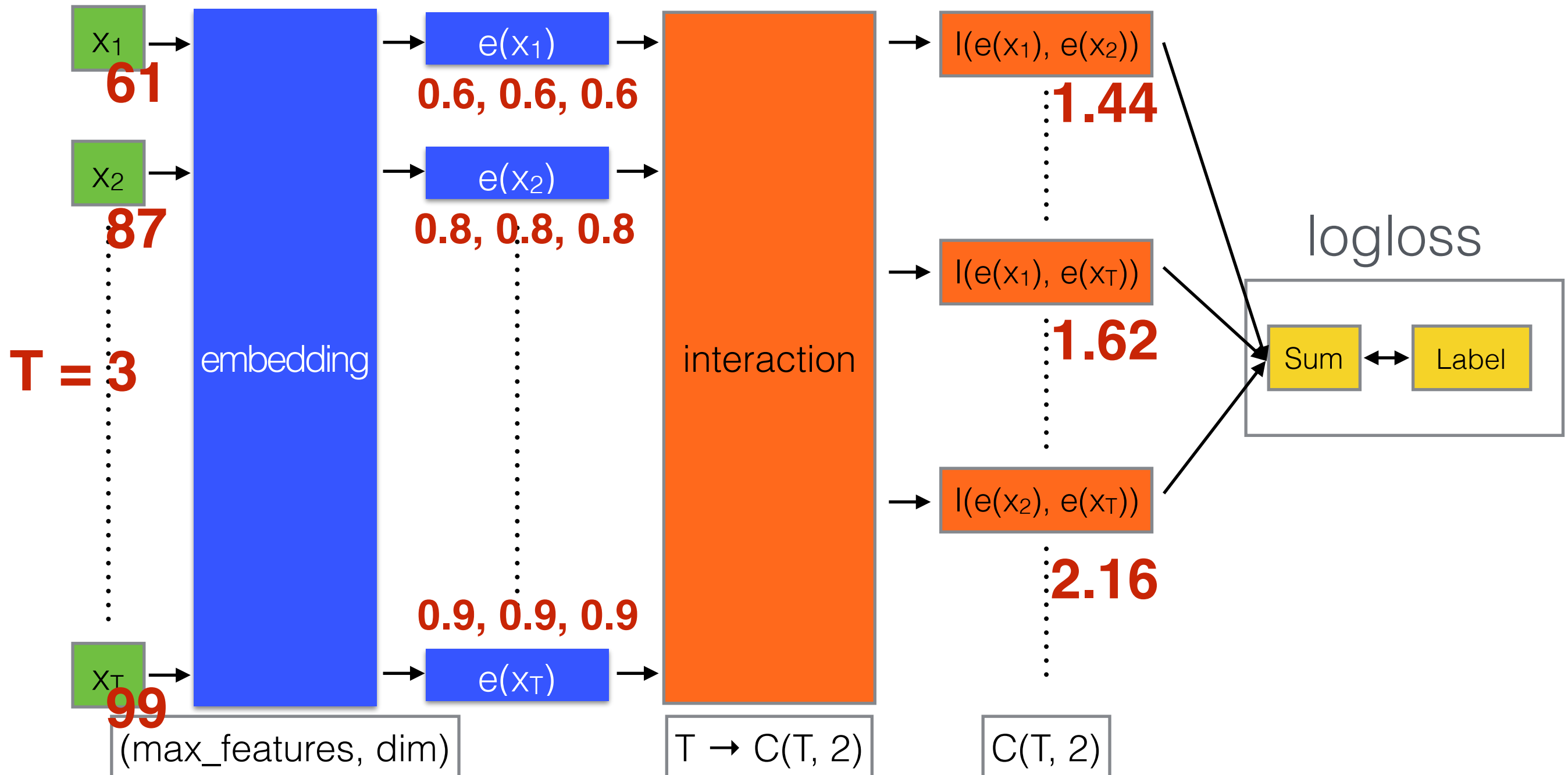
- Each “interaction” has different importance
- Sum \rightarrow Weighted Sum

Weighted FM

example

dim = 3

$C(3, 2) = 3$

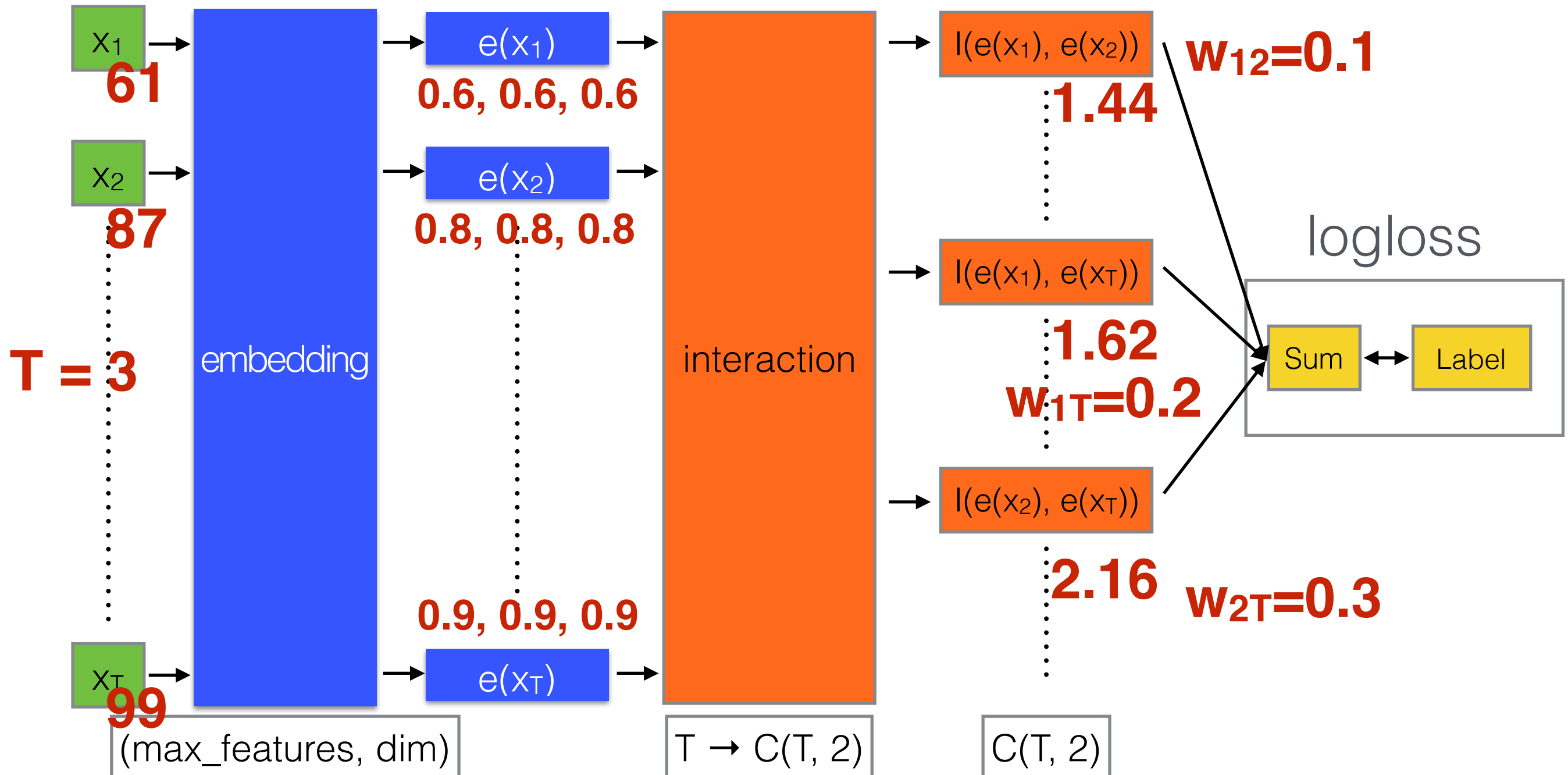


Weighted FM

example

dim = 3

$C(3, 2) = 3$

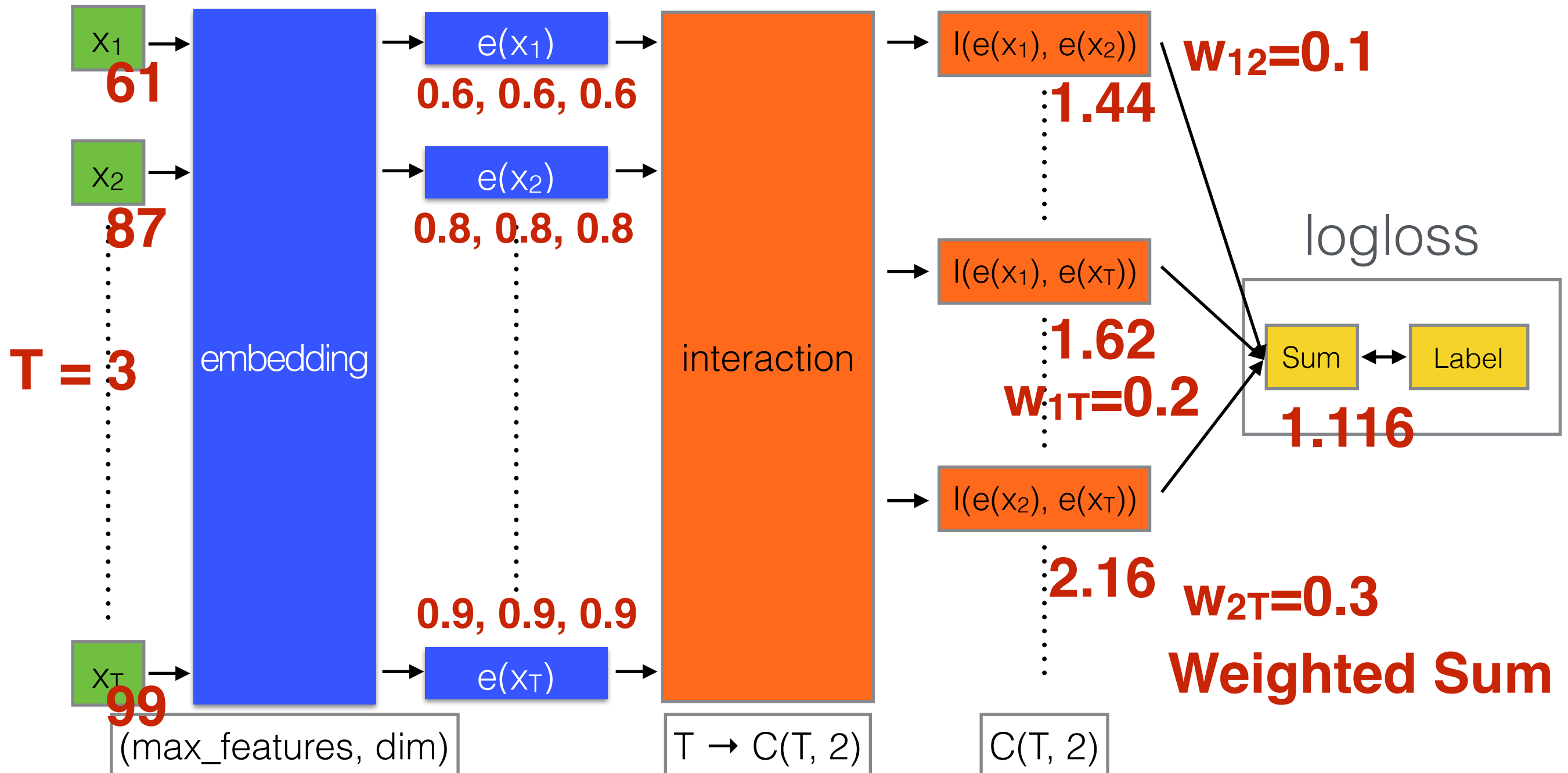


Weighted FM

example

dim = 3

$C(3, 2) = 3$



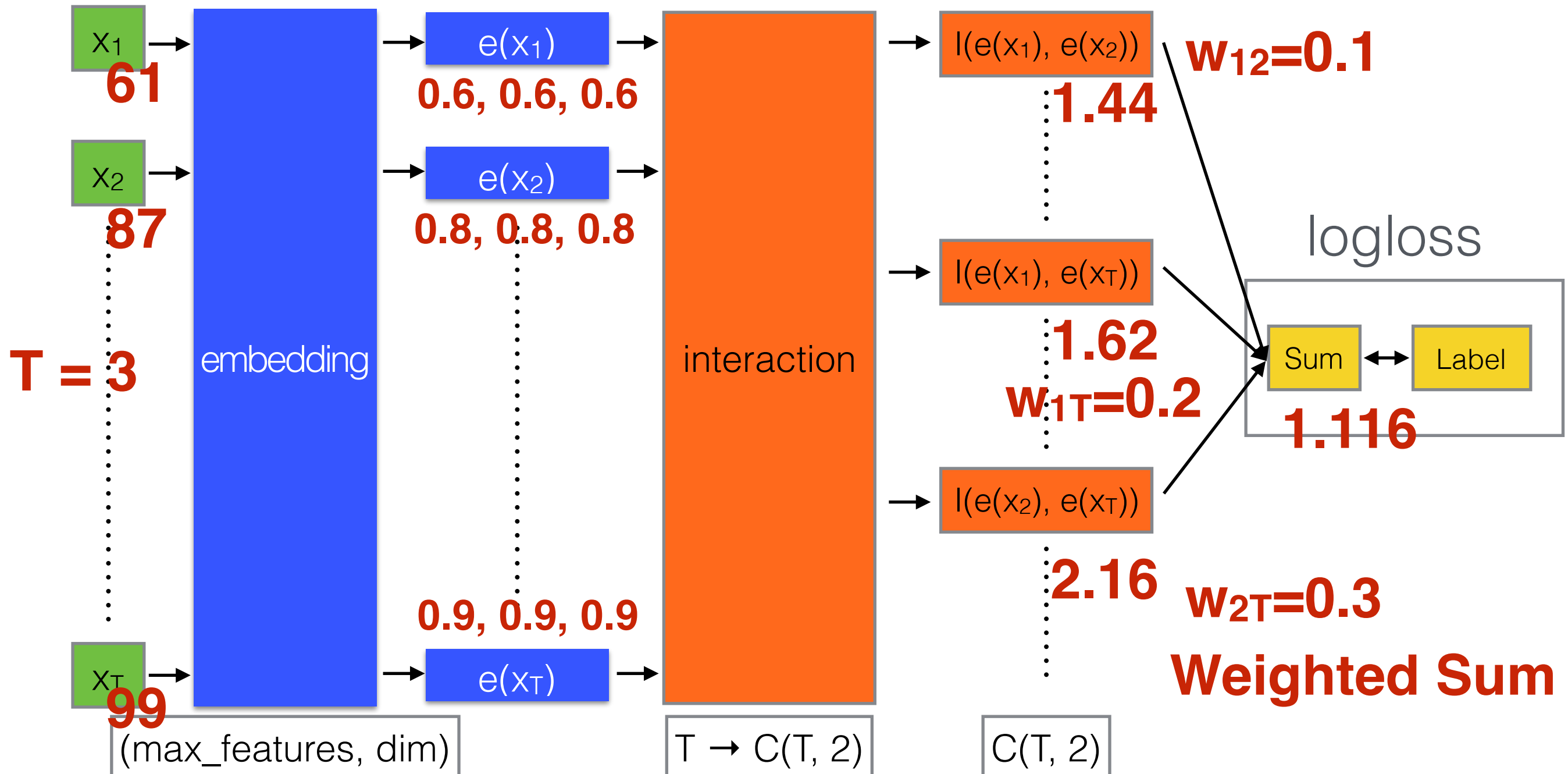
Weighted FM

more weights?

example

dim = 3

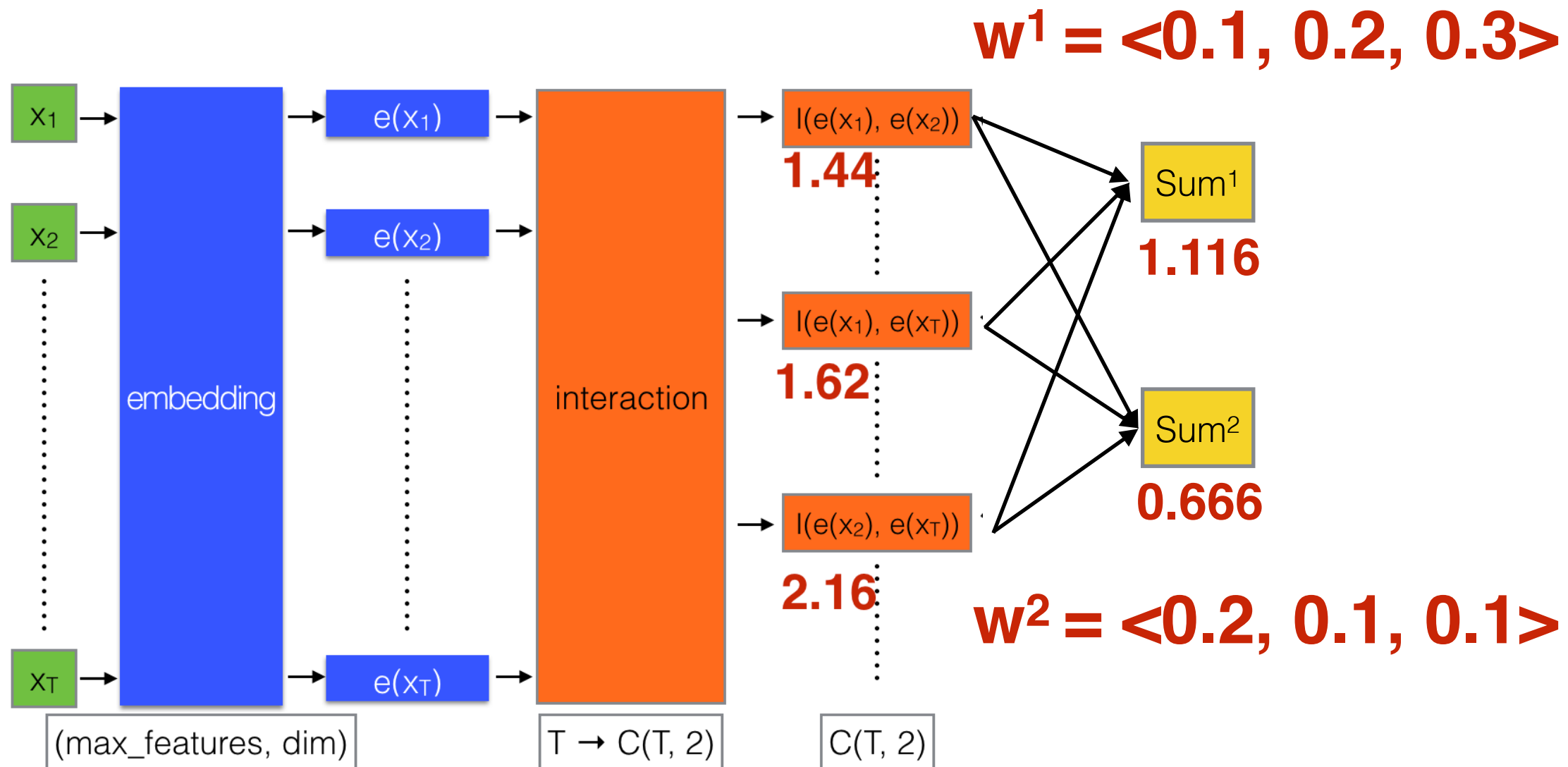
$C(3, 2) = 3$



Weighted FM

more weights?

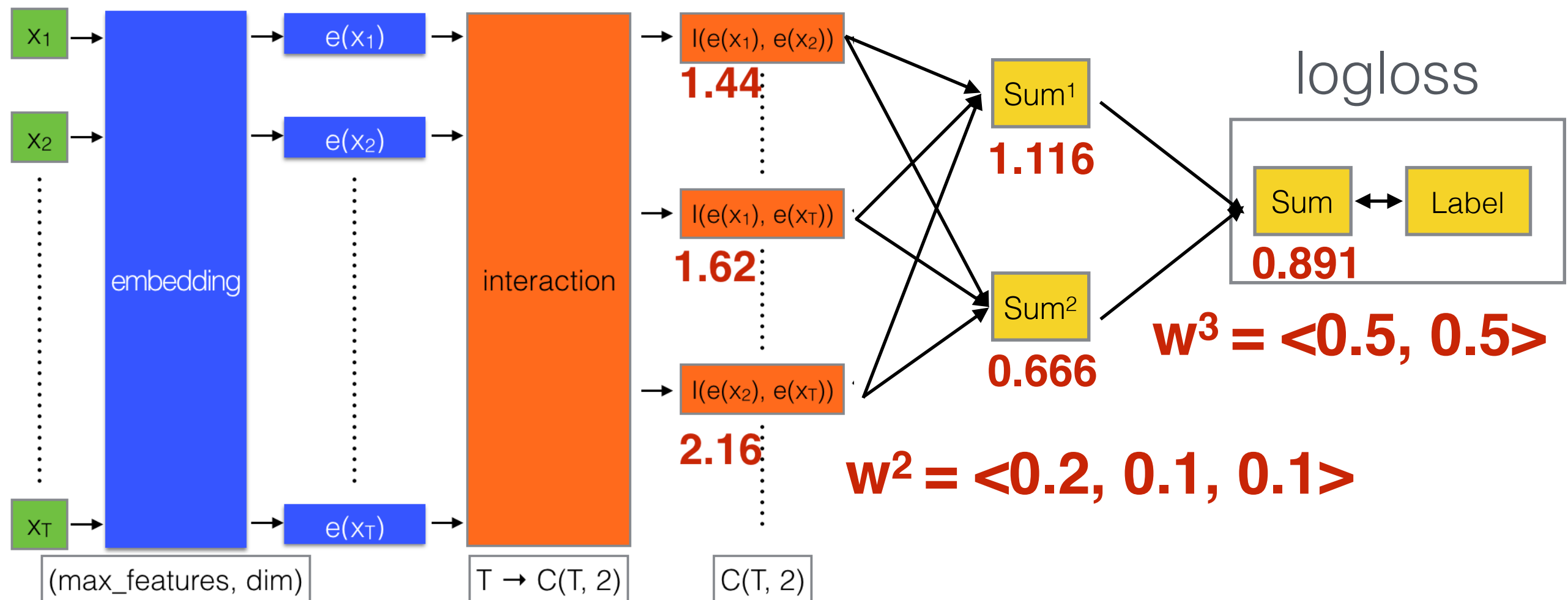
example



FM-NN

Neuron Network?

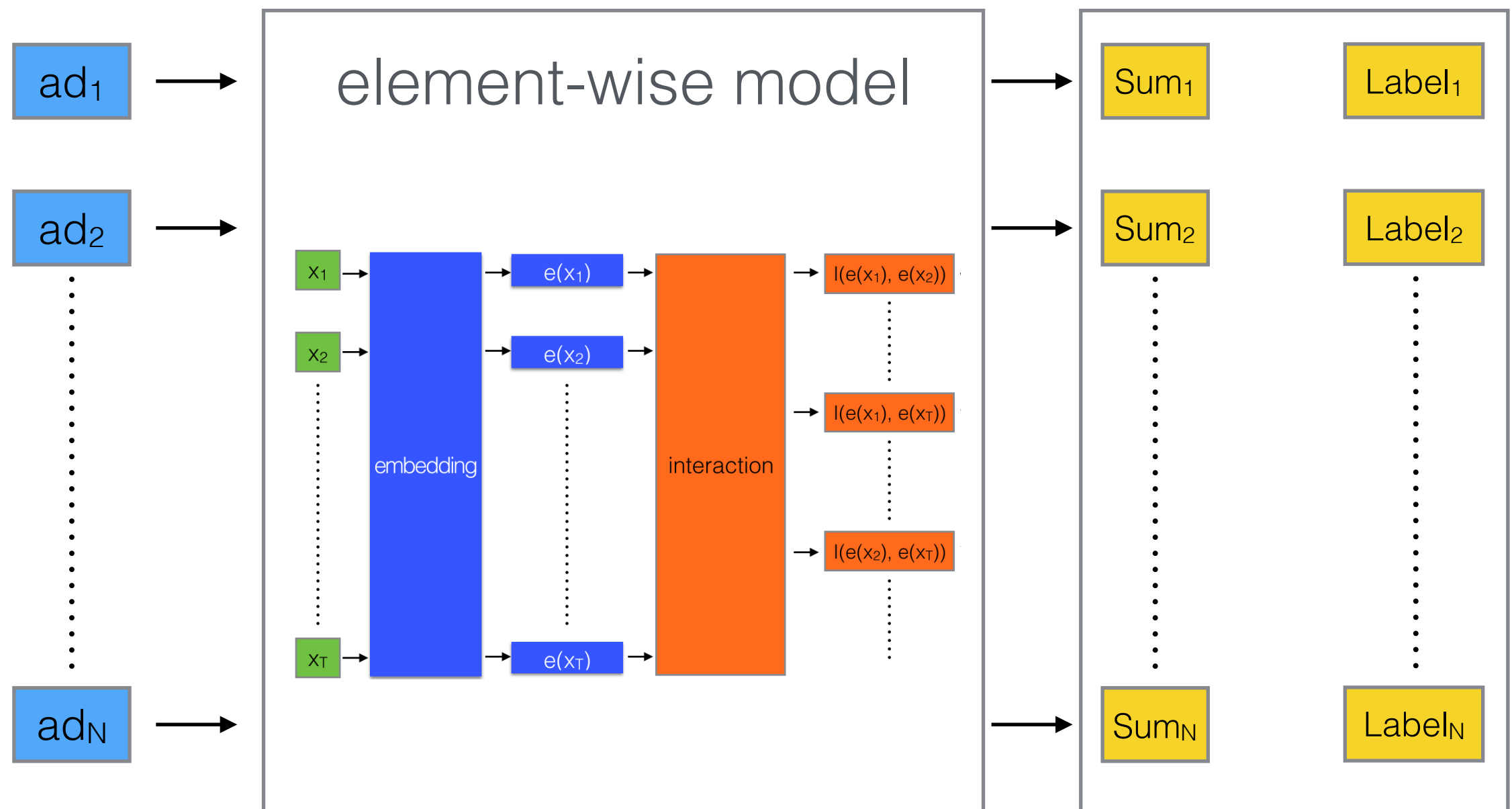
example



List-wise Framework

- Consider the other ads
 - **like** or not (element-wise) → **prefer** which (list-wise)
- a display_id has N ads ($N_{\max}=12$)

List-wise Framework

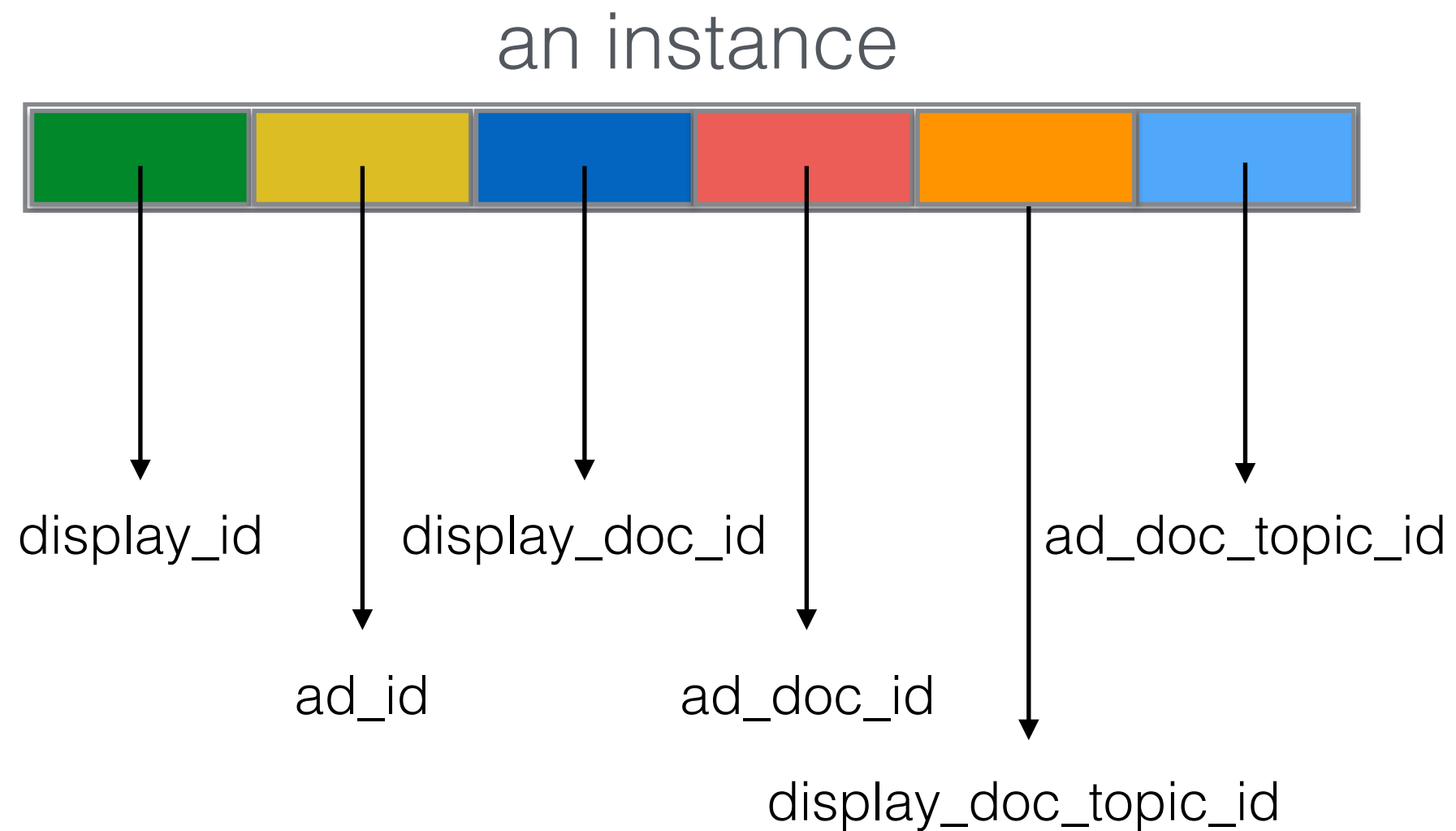


Softmax
& cross entropy
or other cost functions

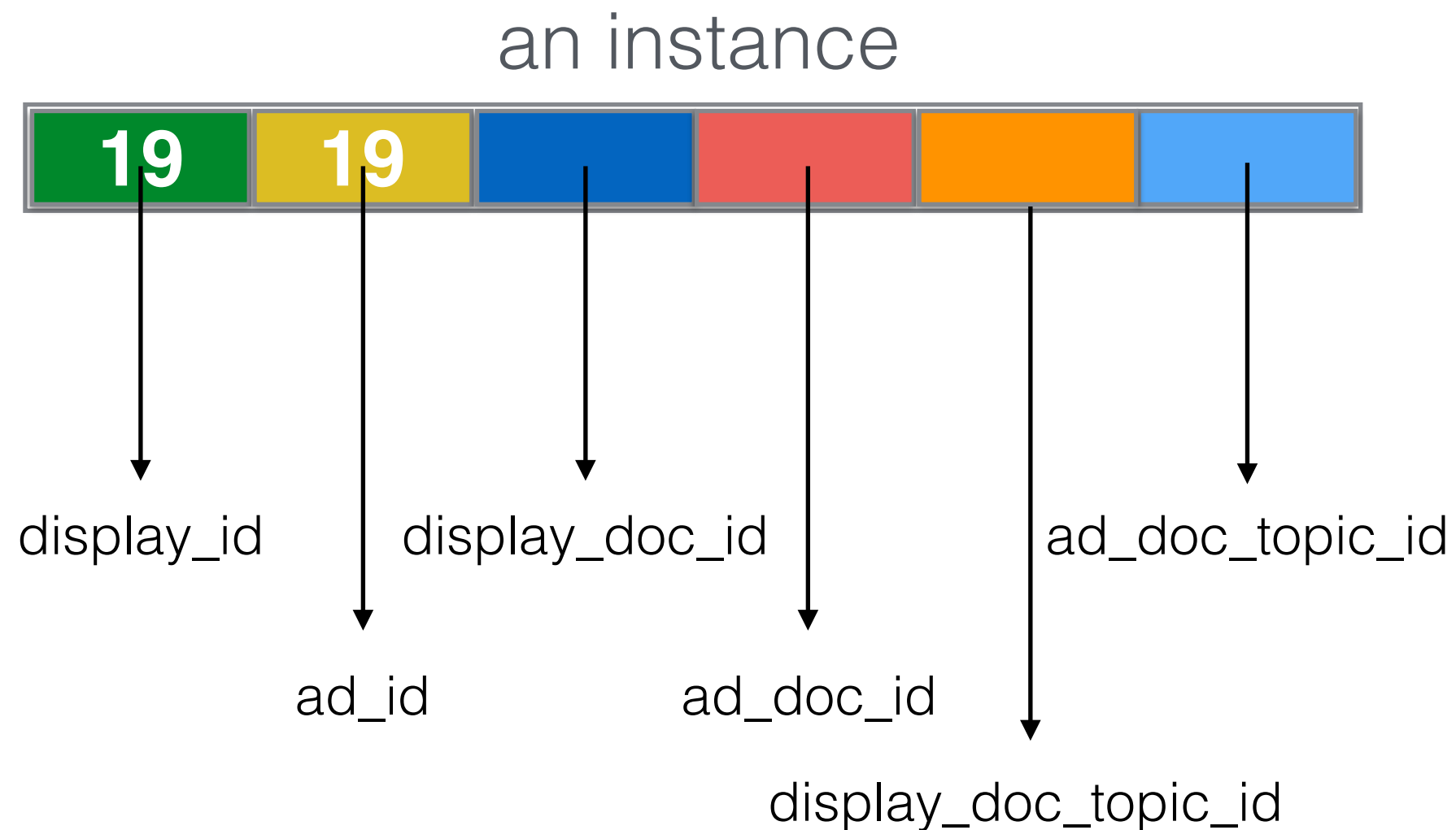
Feature Extraction

- Categorical Features
- Hash Trick

Categorical Features



Categorical Features



**If display_id = 19 and ad_id = 19?
=> the same embedding?**

Hash Trick

- Hash string: “\$category_\$value”

example

display_id=19

hash(“display_id_19”) = 61

ad_id=19

hash(“ad_id_19”) = 37

display_doc_id=6

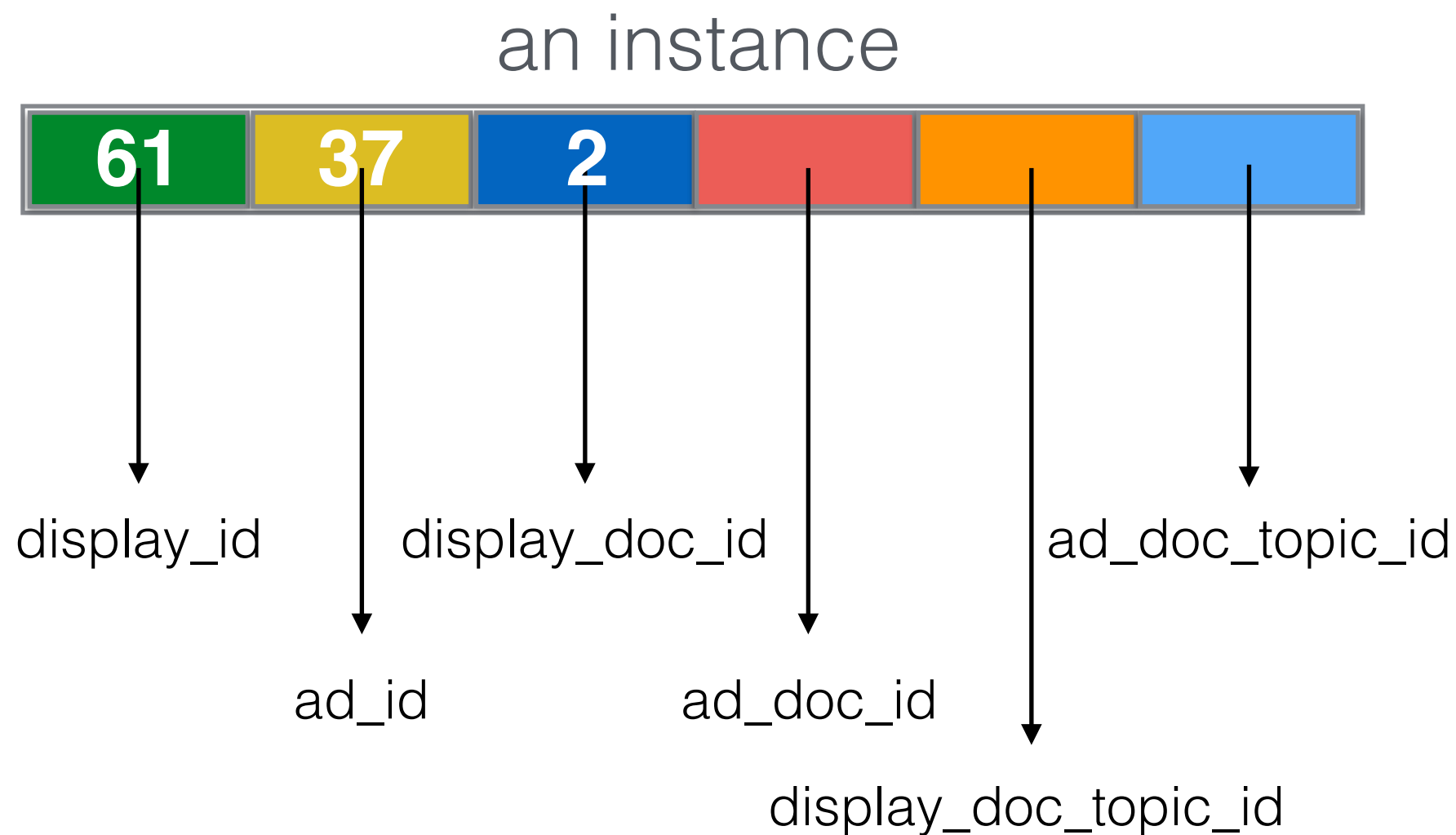
hash(“display_doc_id_6”) = 2

Hash Trick

`hash("display_id_19") = 61`

`hash("ad_id_19") = 37`

`hash("display_doc_id_6") = 2`



Evaluation

- Validation Set
- Evaluation Function

Validation Set

UTC-4

training data

2016/6/14 ~ 2016/6/26

testing data

2016/6/14 ~ 2016/6/26 2016/6/27 ~ 2016/6/28

almost 1:1

split data with the distribution

split training data

2016/6/14 ~ 2016/6/24

62×10^6

validation data

2016/6/14 ~ 2016/6/24 2016/6/25 ~ 2016/6/26

24×10^6

Evaluation Function

- leaderboard: MAP@12
 - Mean Average Precision at 12
- MRR = MAP when $N_{\text{relevant}} = 1$
 - Mean Reciprocal Rank

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

Evaluation Function

- leaderboard: MAP@12
 - Mean Average Precision at 12
- MRR = MAP when $N_{\text{relevant}} = 1$
- Mean Reciprocal Rank

rank	1	2	3
label	0	0	1
MRR		1/3	

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

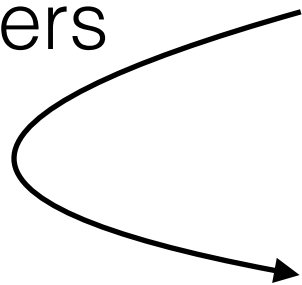
Experiments

	MAP@12 _{Kaggle}
element-wise + 7 features	0.66153
element-wise + 10 features	0.67309
list-wise + 10 features	0.67419

different frameworks and features

Experiments

+ “leak” feature
and others



	MAP@12 _{Kaggle}
element-wise + 7 features	0.66153
element-wise + 10 features	0.67309
list-wise + 10 features	0.67419

different frameworks and features

Experiments

		MAP@12 _{Kaggle}
+“leak” feature and others	element-wise + 7 features	0.66153
	element-wise + 10 features	0.67309
	list-wise + 10 features	0.67419

+list-wise framework
+weighted sum

different frameworks and features

Experiments

	MAP@12 _{Kaggle}
+ list-wise + 10 features + dim 4	0.67419
+ list-wise + 10 features + dim 8	0.67652
+ list-wise + 10 features + dim 64	0.67816

different embedding dimensions

Any Questions?

