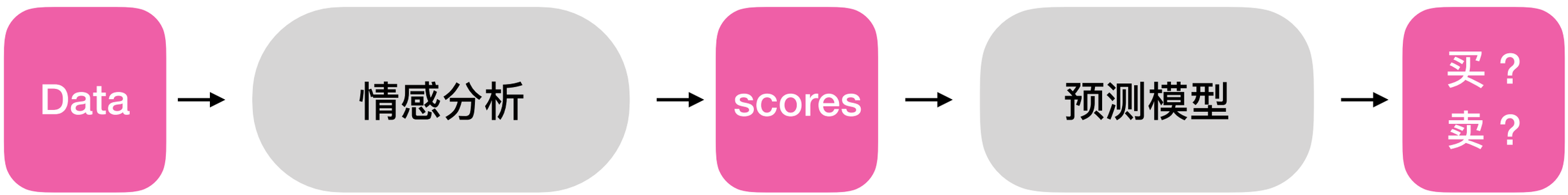
模型架构

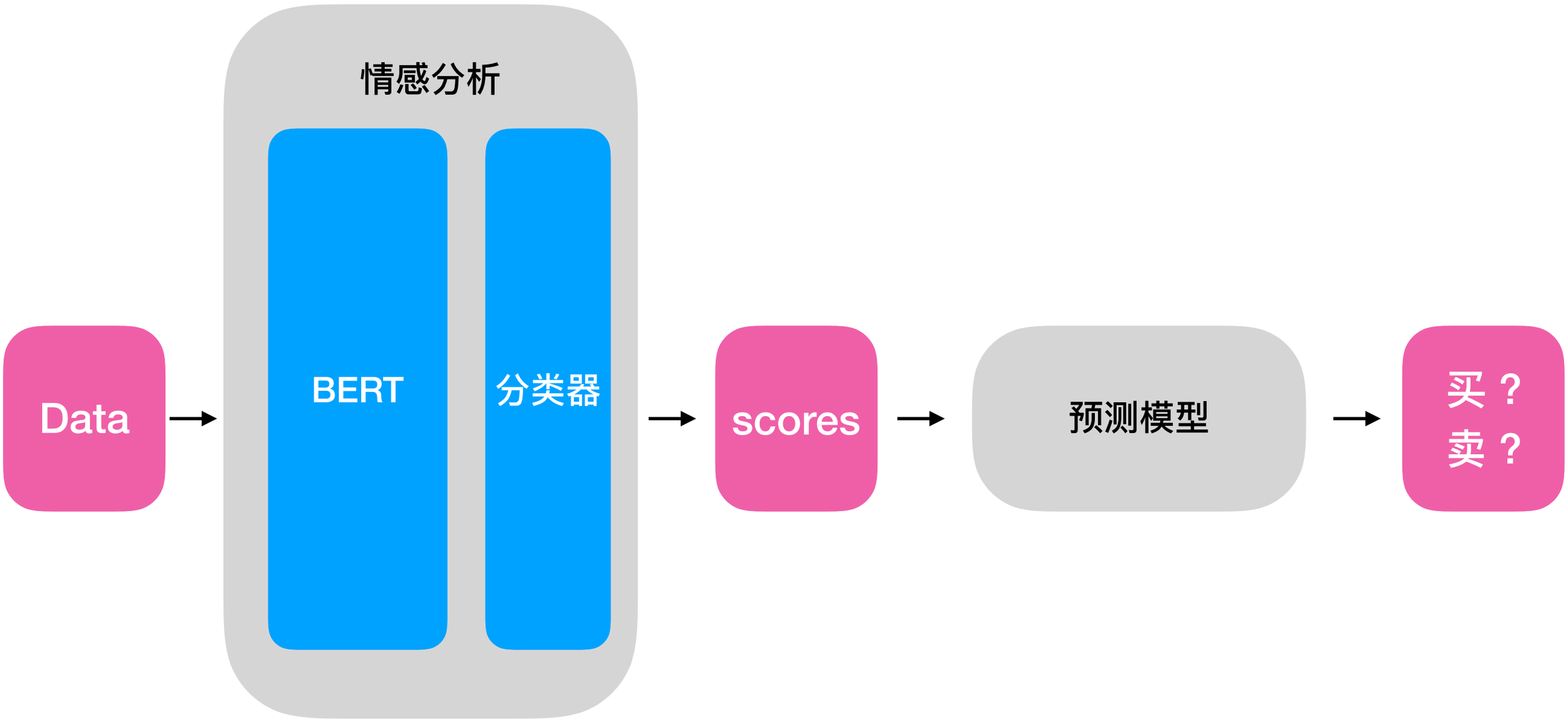


数据来源

Twitter HashTag
e.g. $AAL

Yahoo Finance

模型架构



BERT

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

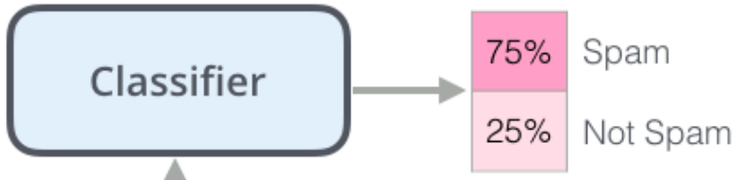
Model:

Dataset:

Objective: Predict the masked word (langauge modeling)

2 - Supervised training on a specific task with a labeled dataset.

Supervised Learning Step



Model: (pre-trained in step #1)

Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

BERT

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	**Average** -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

BERT

- 词向量模型
  - word2vec
  - ELMo

BERT

- word2vec
  - 上下文无关
  - CBOW/Skip-Gram

BERT

- word2vec
  - 上下文无关
  - CBOW/Skip-Gram
- ELMo
  - 上下文相关
  - LSTM
    - 并行能力差

BERT

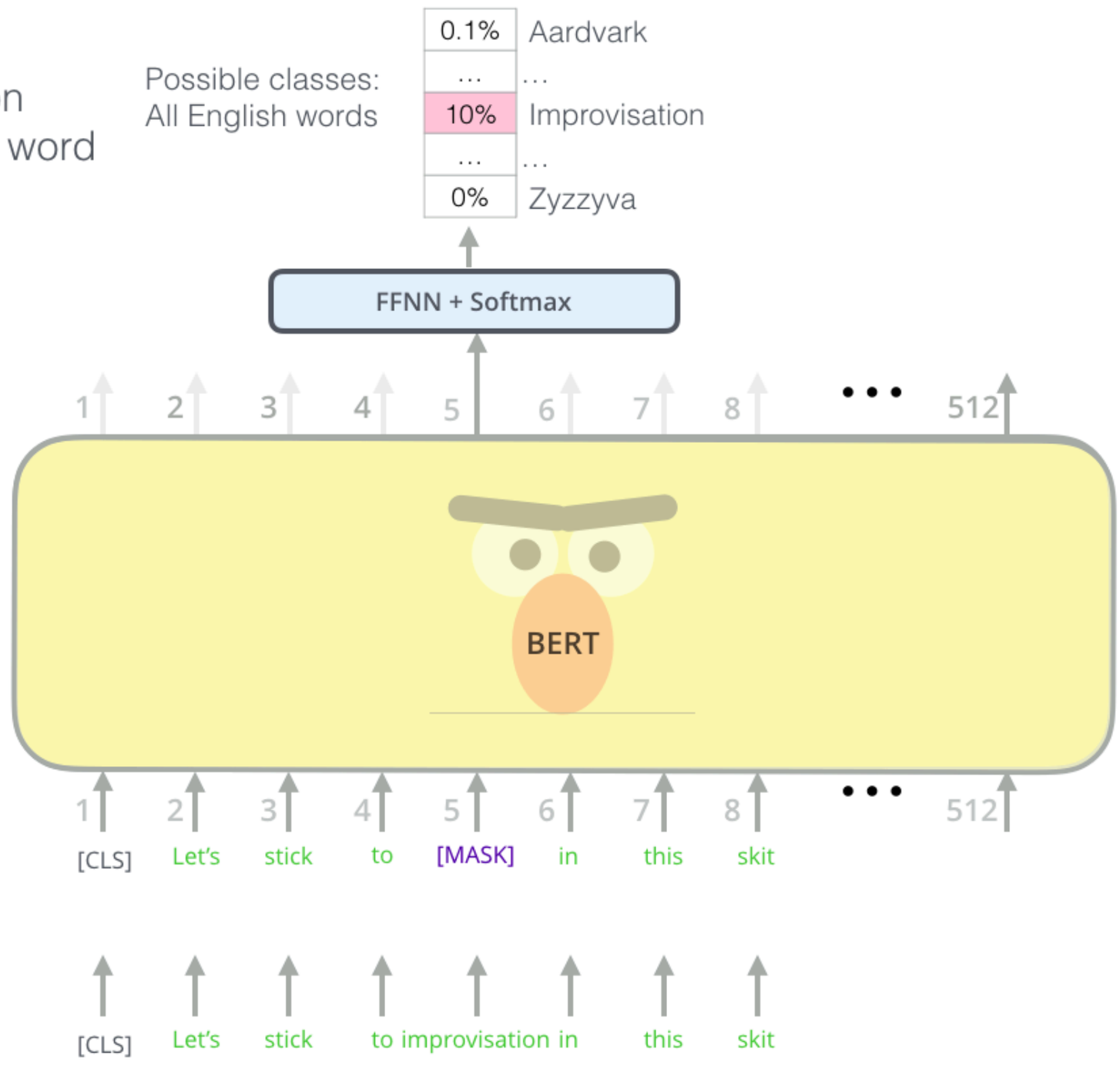
- word2vec
  - 上下文无关
  - CBOW/Skip-Gram
- ELMo
  - 上下文相关
  - LSTM
    - 并行能力差
- BERT
  - 上下文相关
  - Transformer
    - 并行能力好
    - 特征提取能力更强
  - Subwords

BERT - Subwords

- “BERT tokenizer”
- Word
  - [‘bert’, ‘tokenizer’]
- Subword
  - [‘bert’, ‘token’, ‘##izer’]

BERT - Masked Language Model

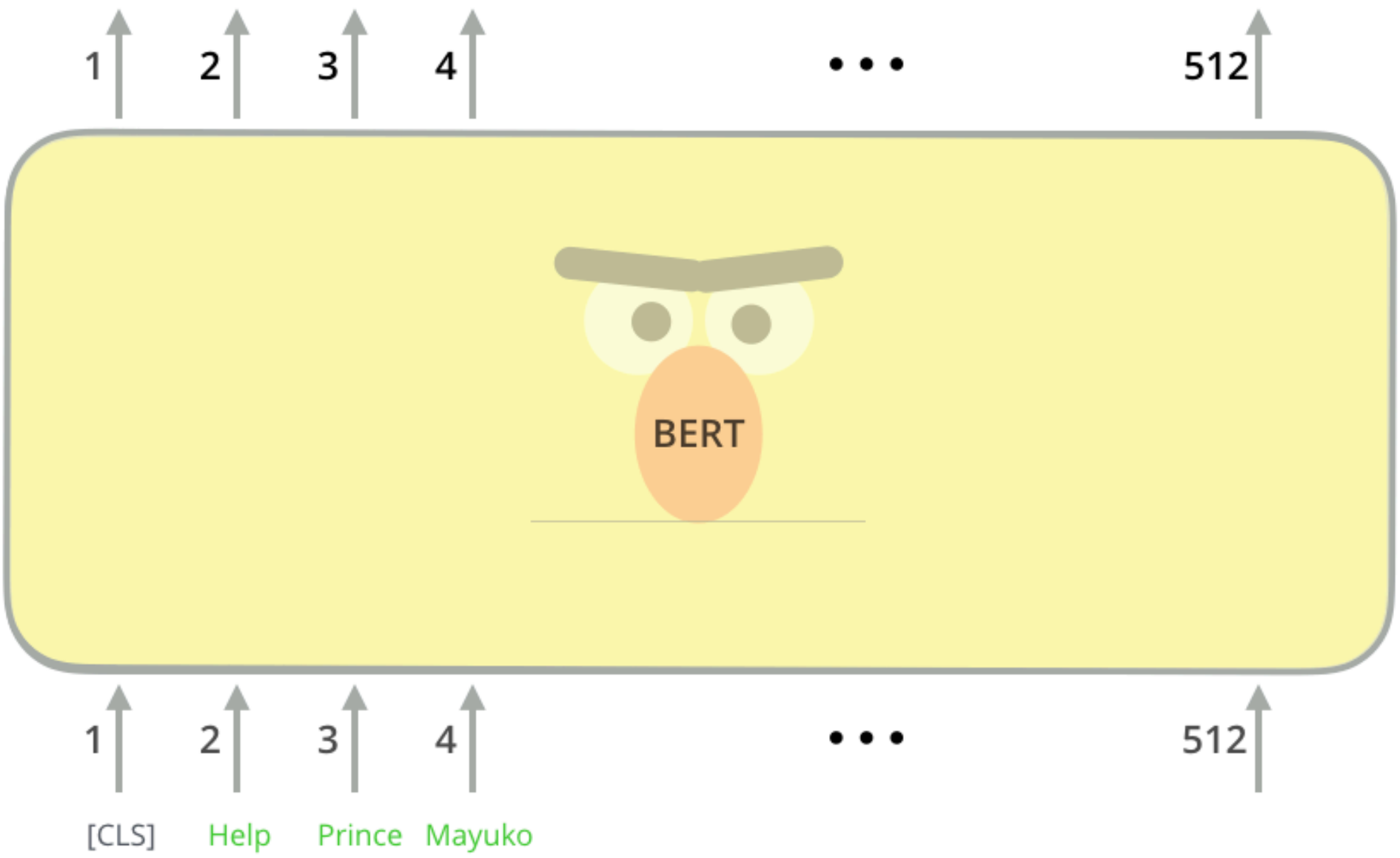
Use the output of the masked word's position to predict the masked word



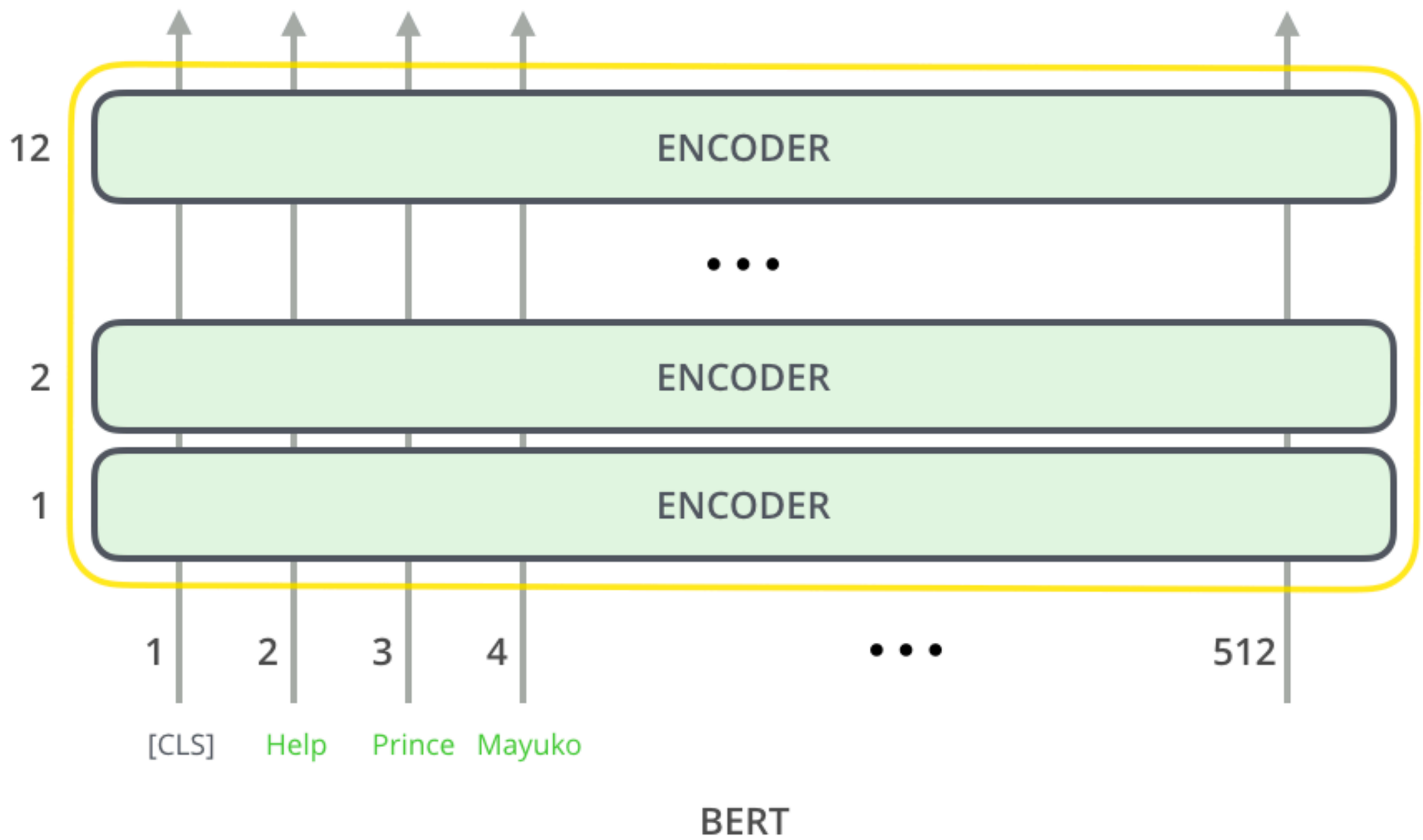
Randomly mask 15% of tokens

Input

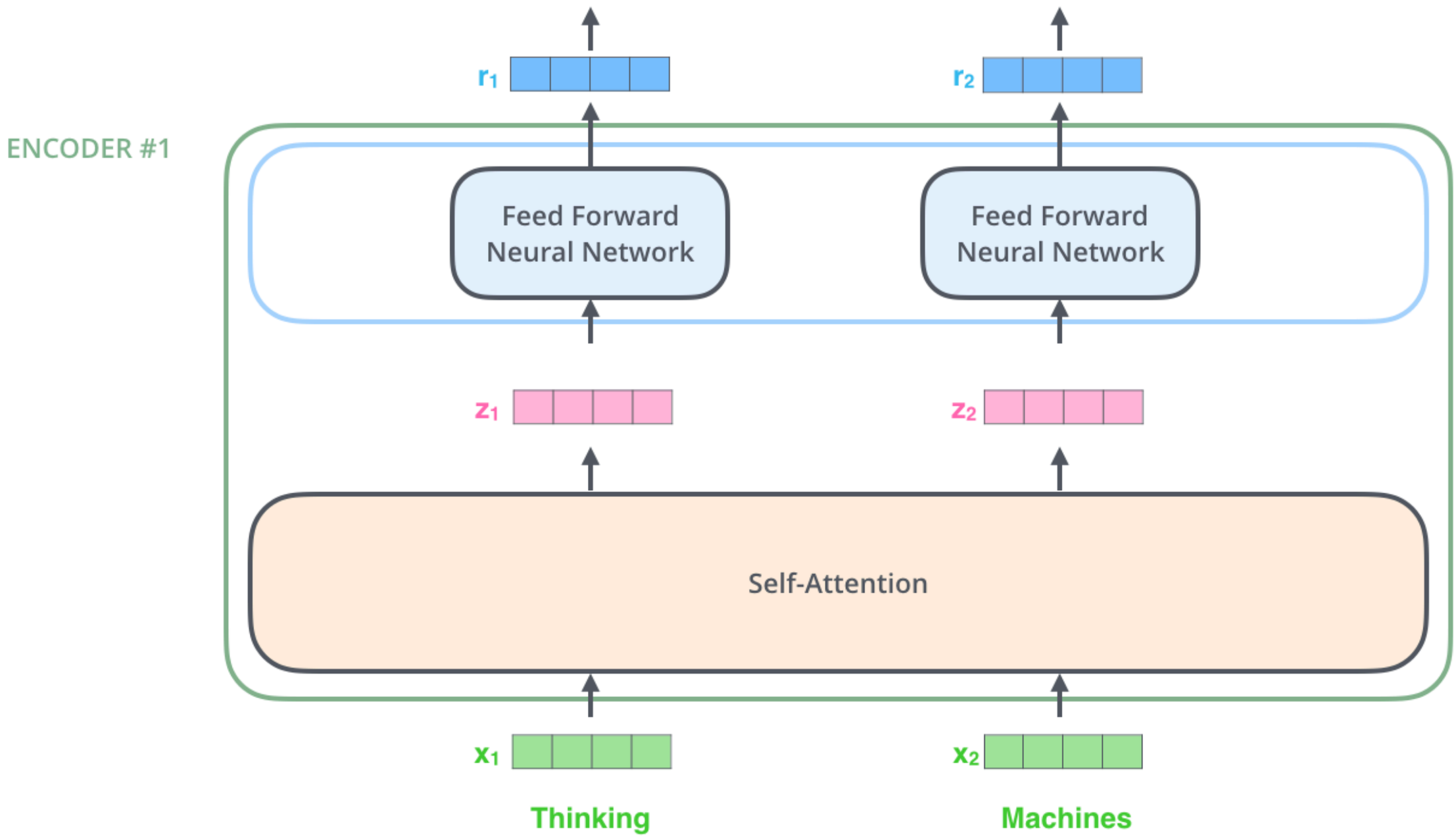
BERT - Architecture



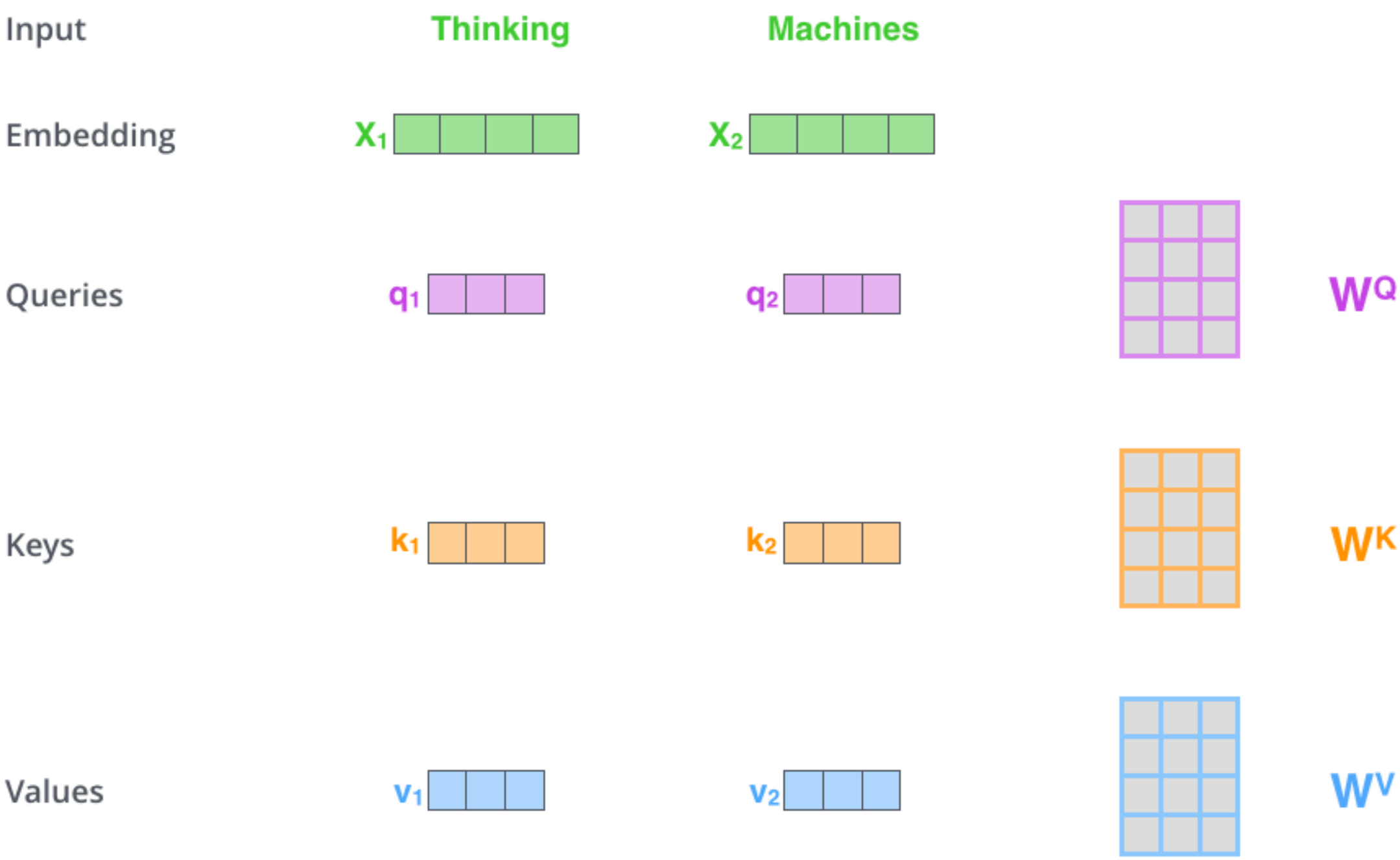
BERT - Architecture



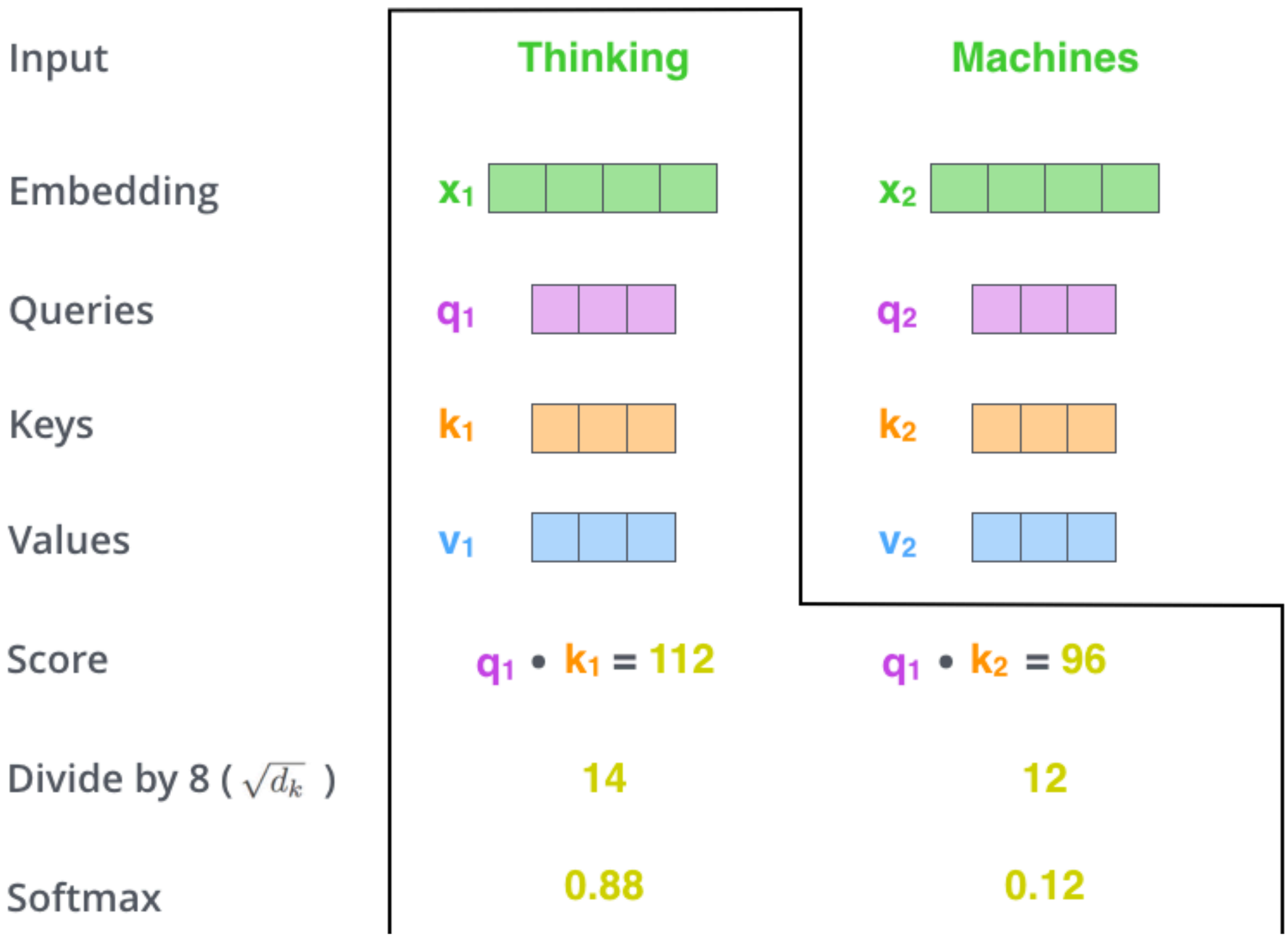
BERT - Architecture



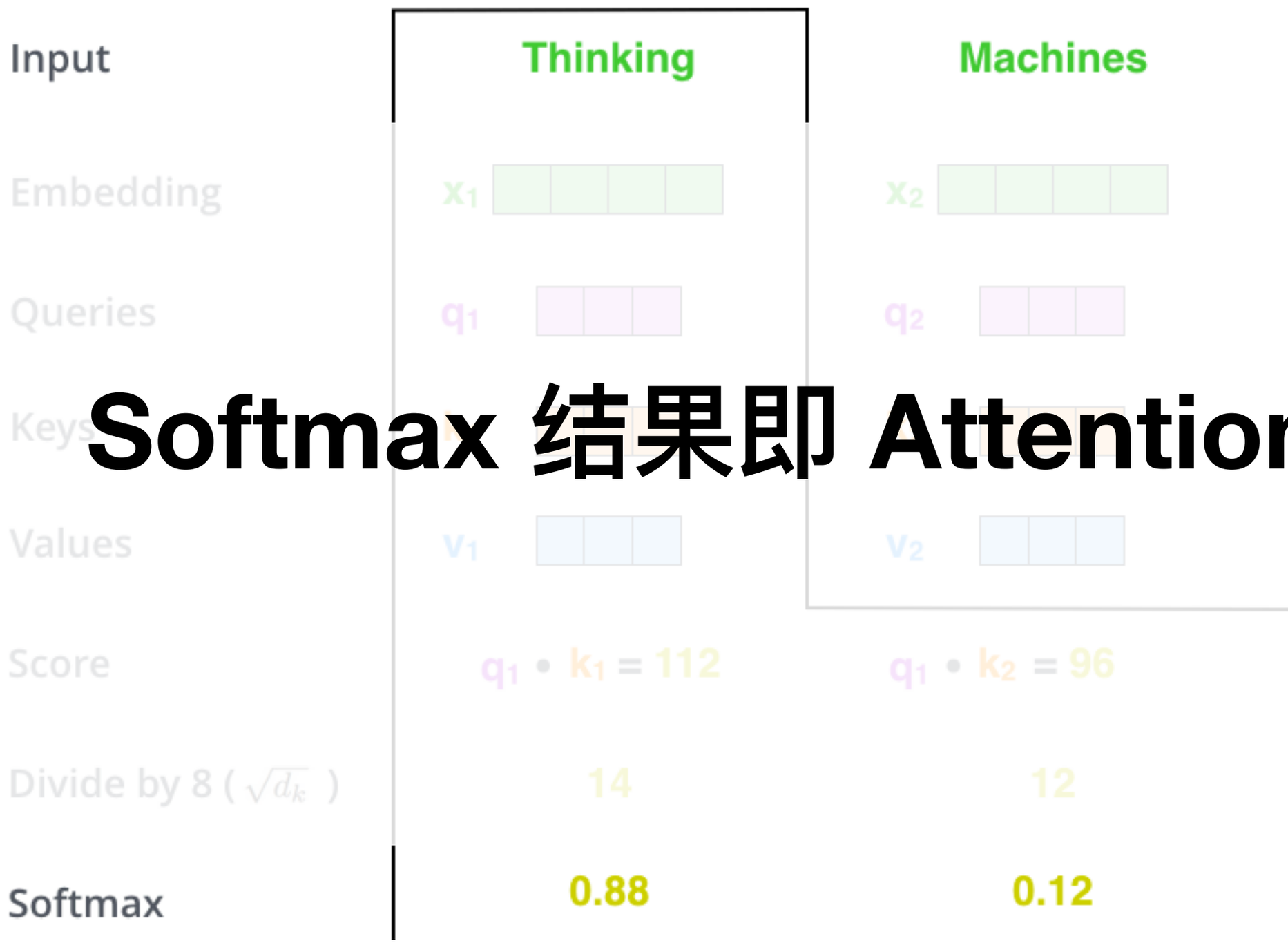
BERT - Self-attention



BERT - Self-attention

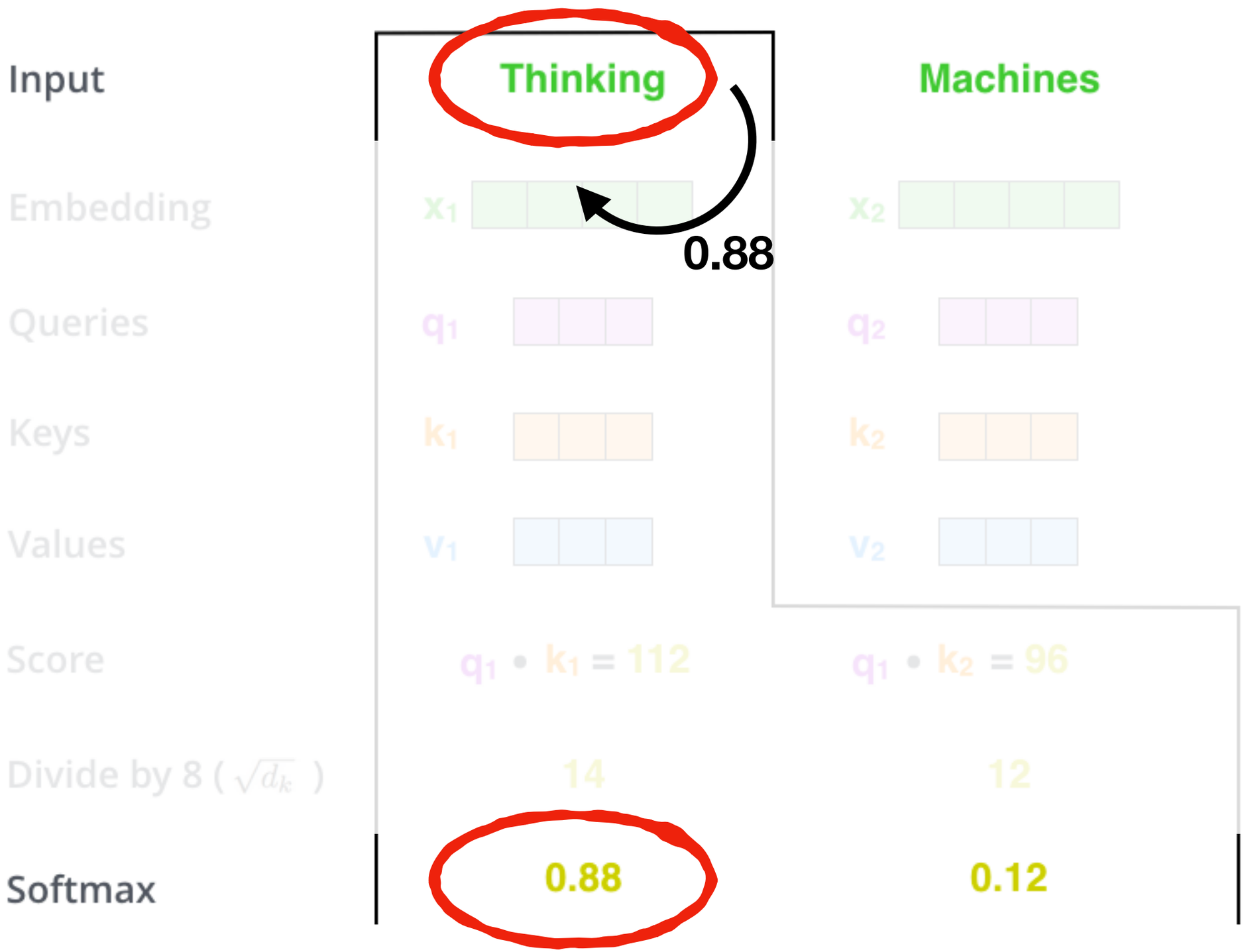


BERT - Self-attention

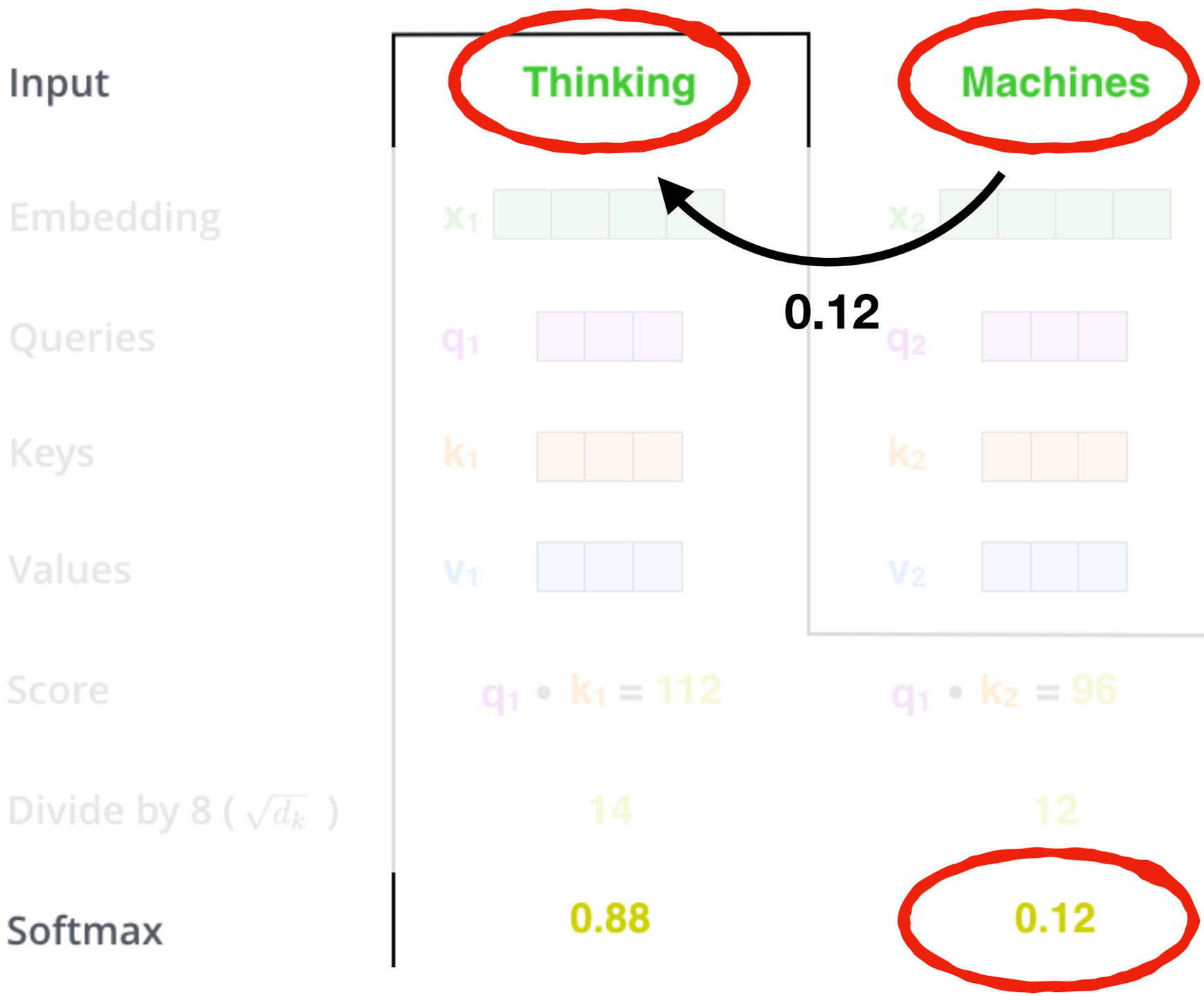


Softmax 结果即 Attention

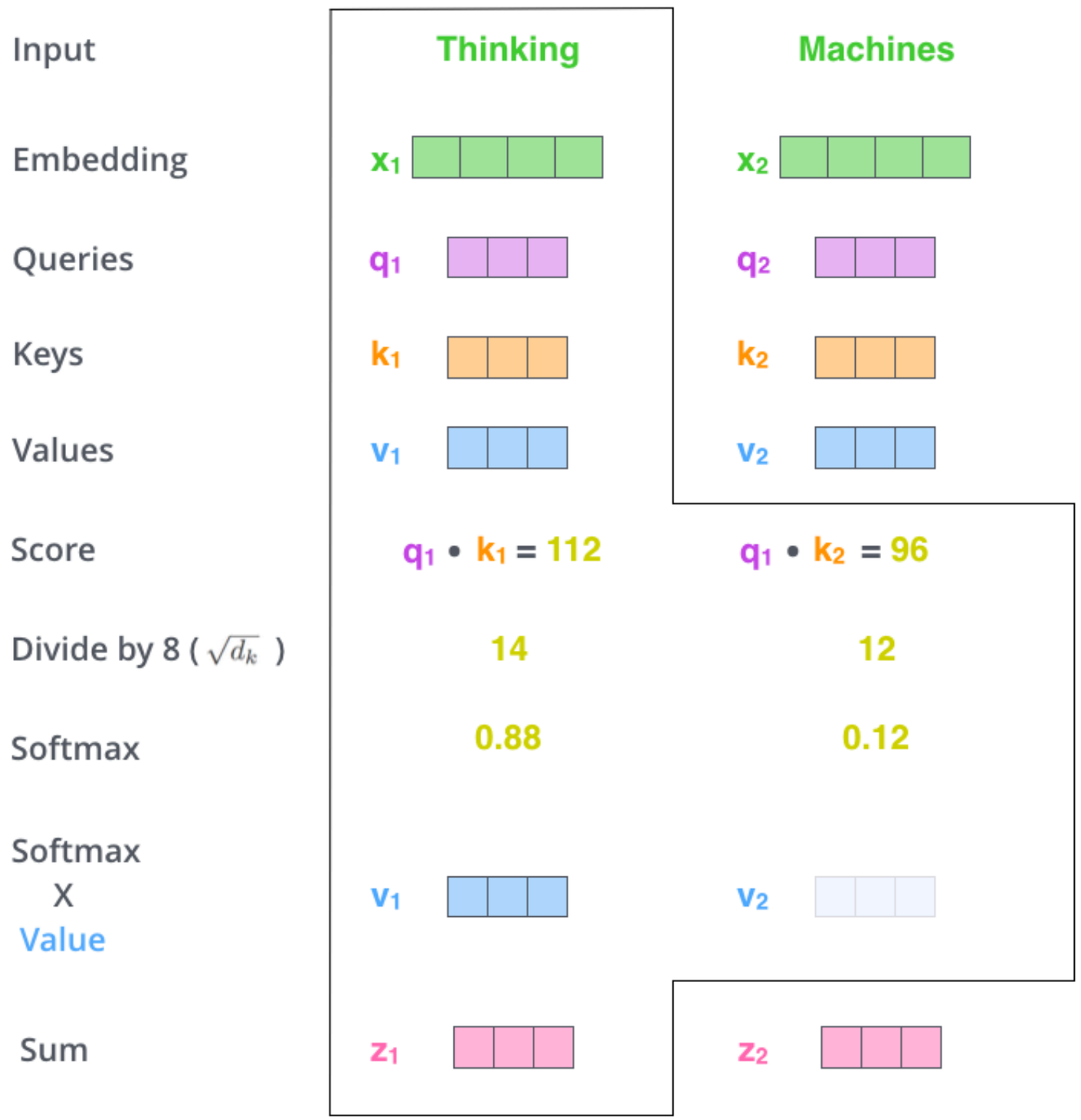
BERT - Self-attention



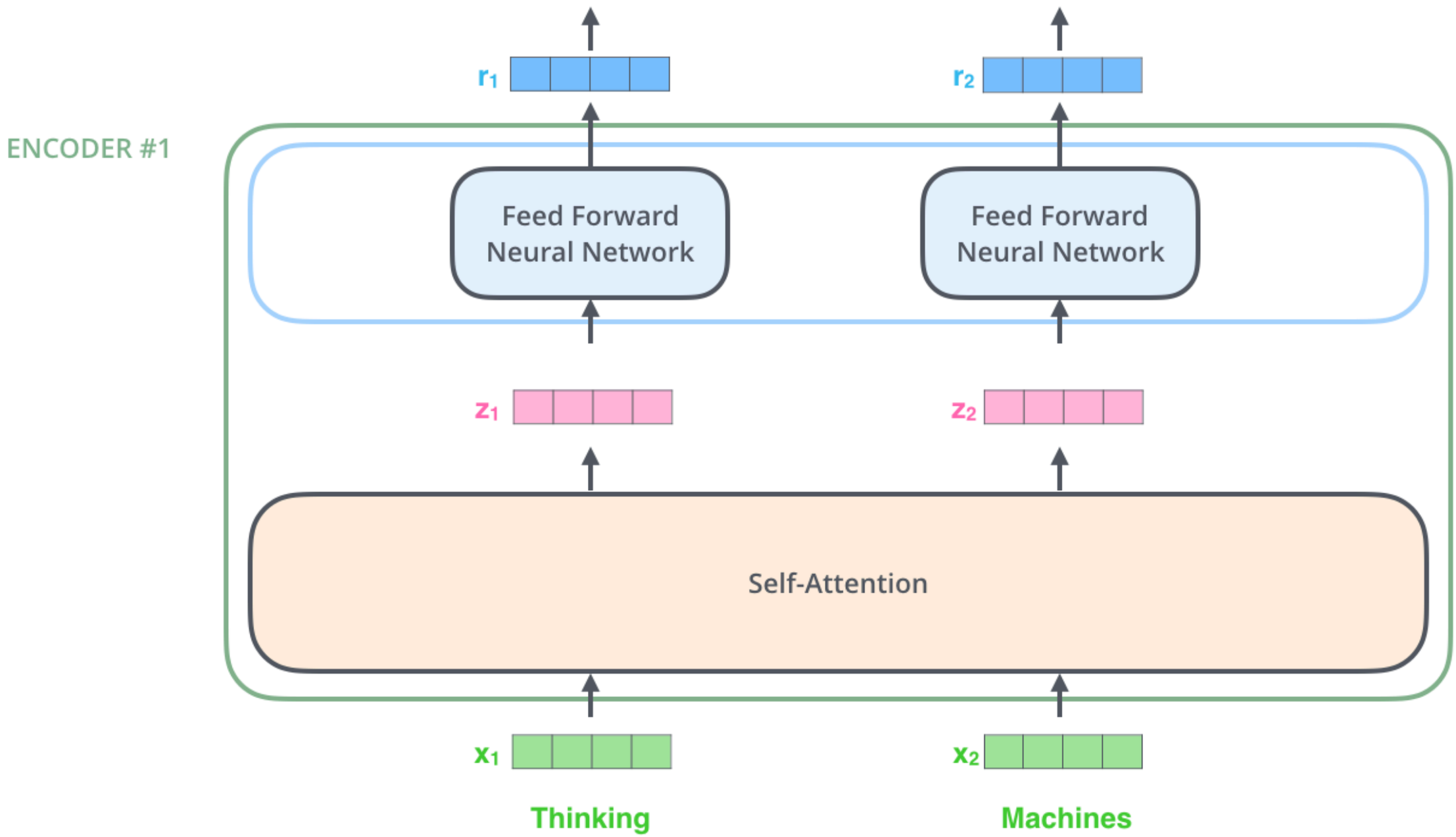
BERT - Self-attention



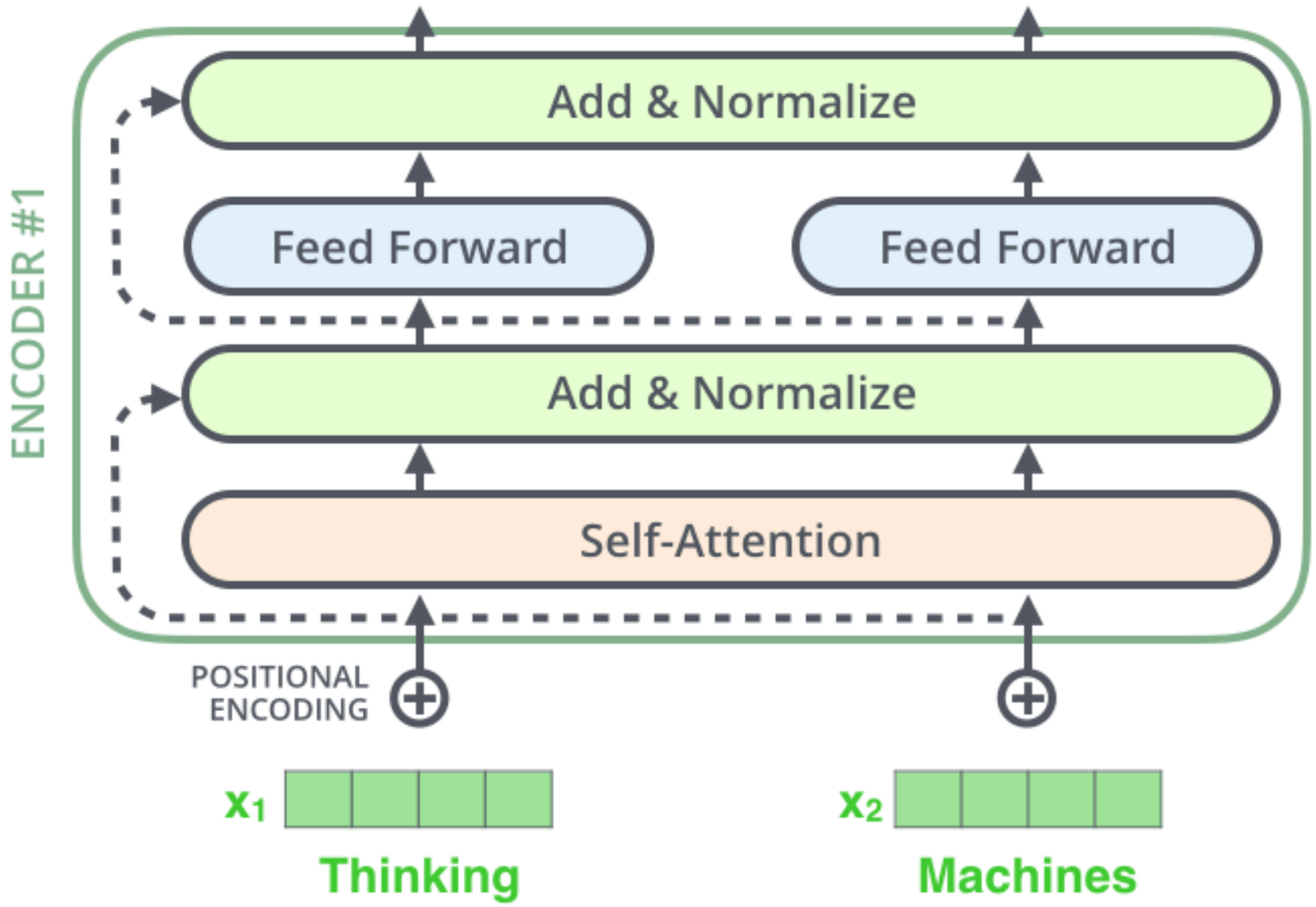
BERT - Self-attention



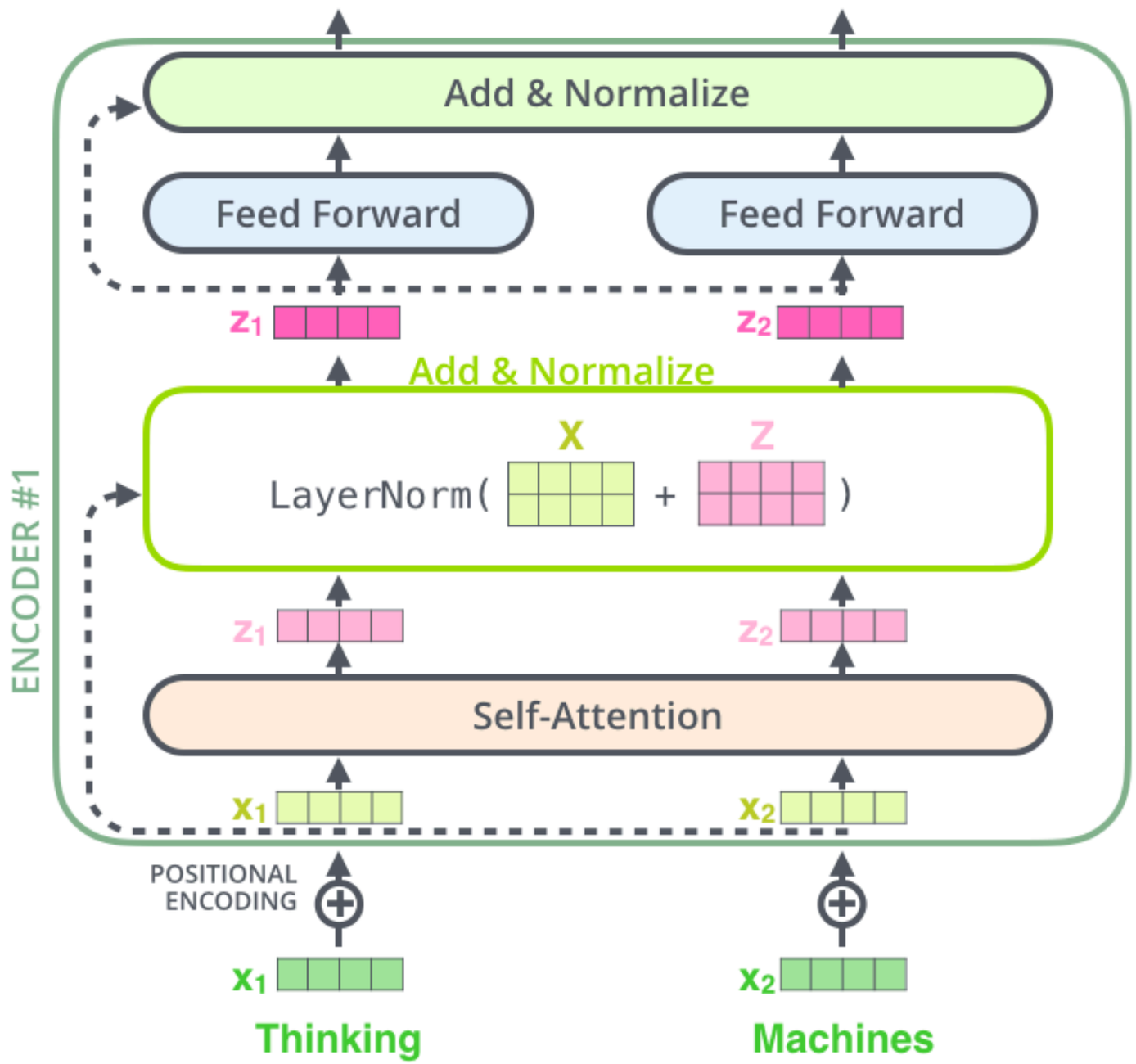
BERT - Self-attention



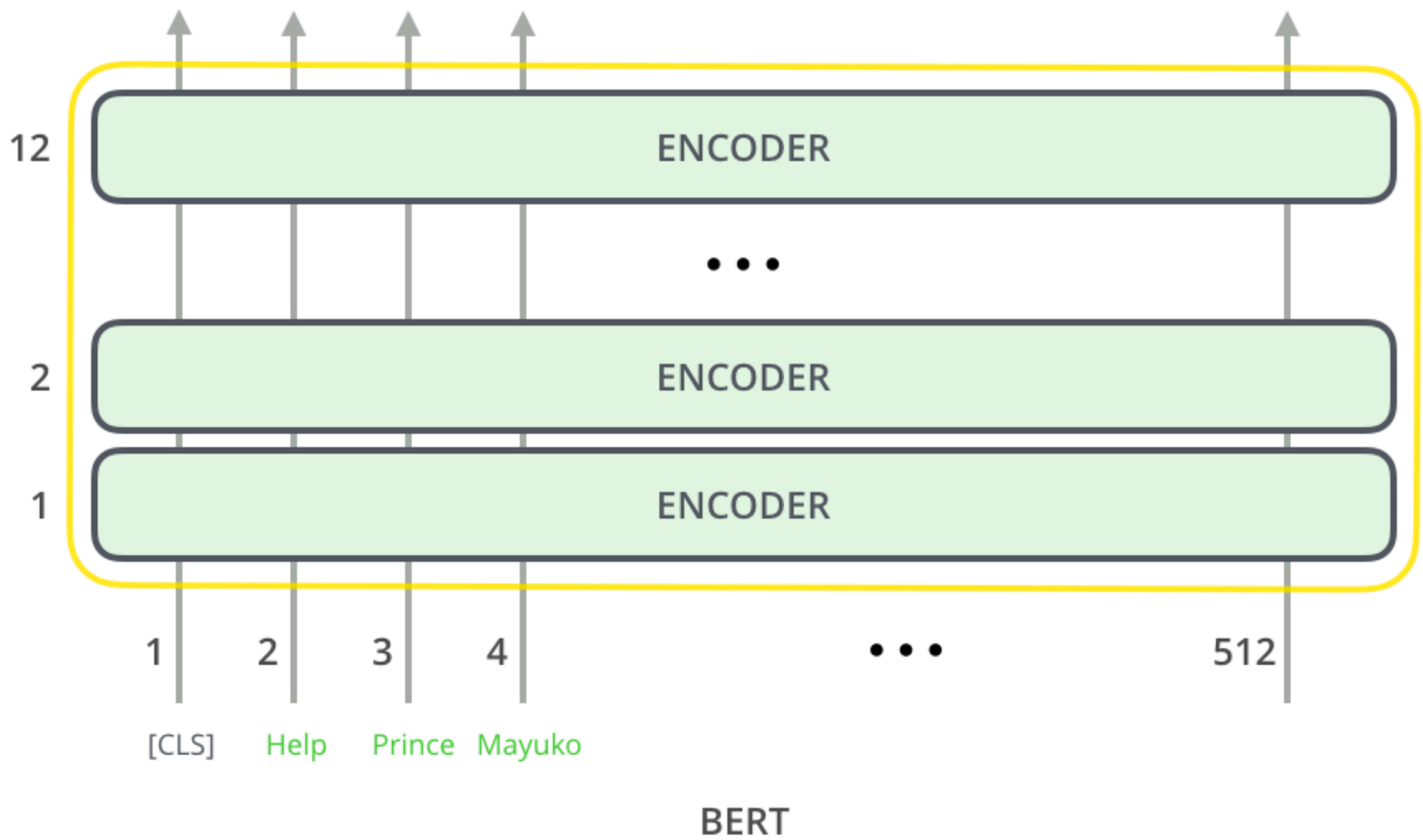
BERT - Self-attention



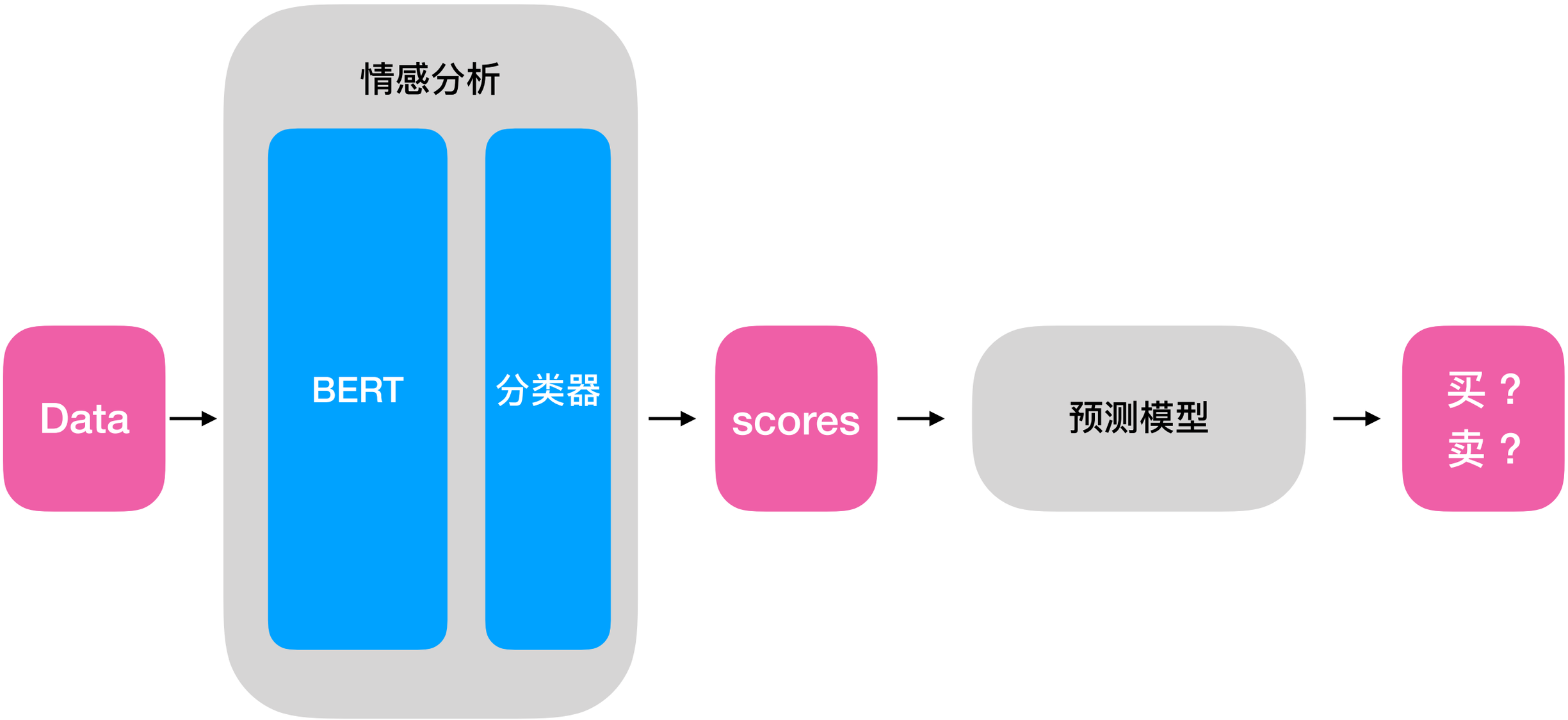
BERT - Self-attention



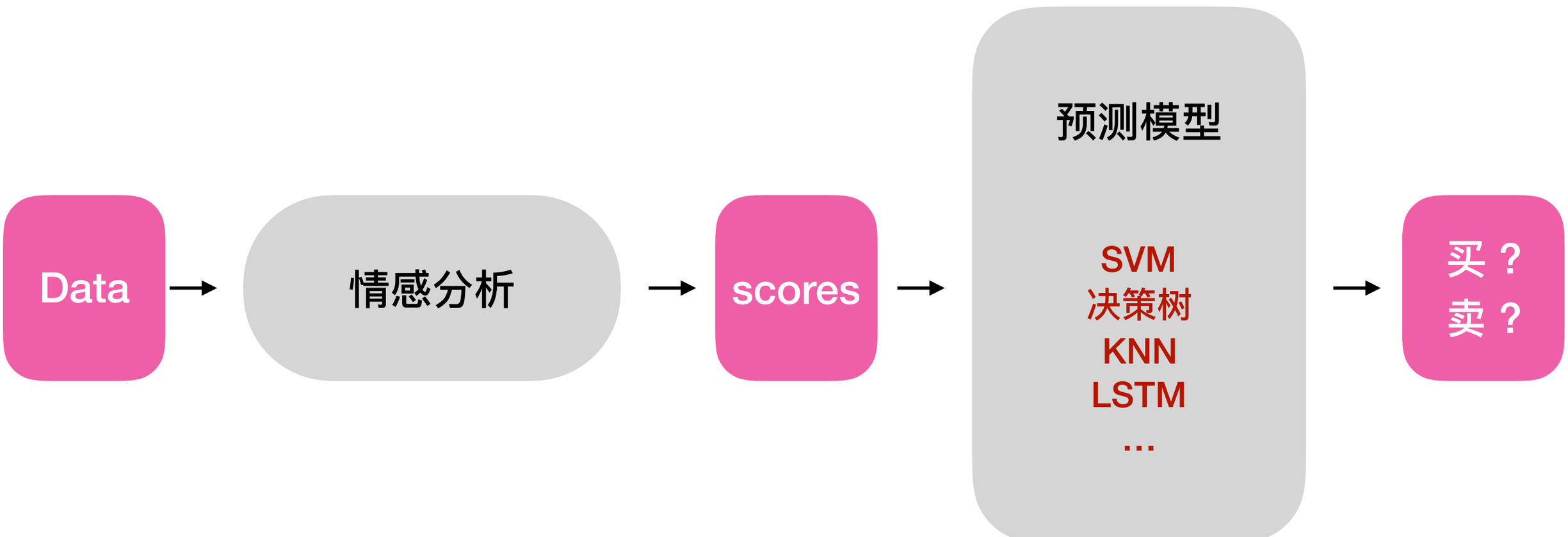
BERT - Architecture



模型架构



模型架构



预测模型

	输入前一天	输入前三天
KNN	0.442381	0.593333
Logistic 回归	0.627619	0.627619
线性核 SVM	0.627619	0.627619
决策树	0.485238	0.510952