



專案主題：
運用NLP技術分析財金新聞

學生：陳冠維 國立清華大學 計財所 碩一

大綱



1

BERT 模型介紹

1. Word Embedding
2. Contextualized Word Embedding
3. Self-Attention & Transformer
4. BERT structure & Training
5. BERT Fine-tune Task

Word Embedding

➤ **1-of-N Encoding:** 將詞轉為向量最簡單的方法

apple = [1 0 0 0 0]

bag = [0 1 0 0 0]

cat = [0 0 1 0 0]

dog = [0 0 0 1 0]

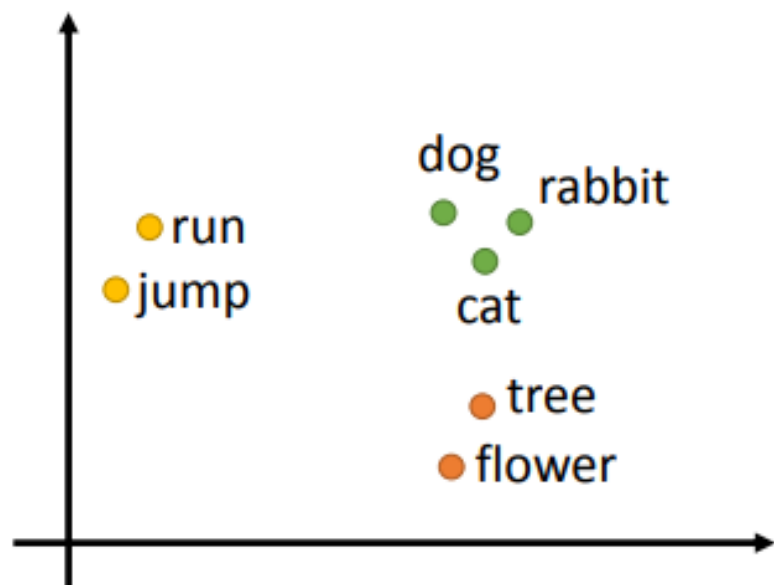
elephant = [0 0 0 0 1]

缺點

- **維度過高:** 假設英文共有10000個單字，則每個單字的向量維度就有10000維
- **詞與詞間不存在關聯性:** 從這些向量中，我們無法看出cat跟dog在某種程度上的關聯

Word Embedding

- **Word Embedding:** 將1-of-N Encoding的高維度向量，透過類神經網路，降維至一個多維度的向量中。並使每個word embedding彼此間具有一定程度的關聯性。



訓練機制

- **非監督式學習:** 訓練word embedding不需要給資料標籤，只需要讓machine讀大量文章
- **透果上下文理解詞與詞的關聯:**

馬英九 宣誓就職.....

蔡英文 宣誓就職.....

Word Embedding (Training)

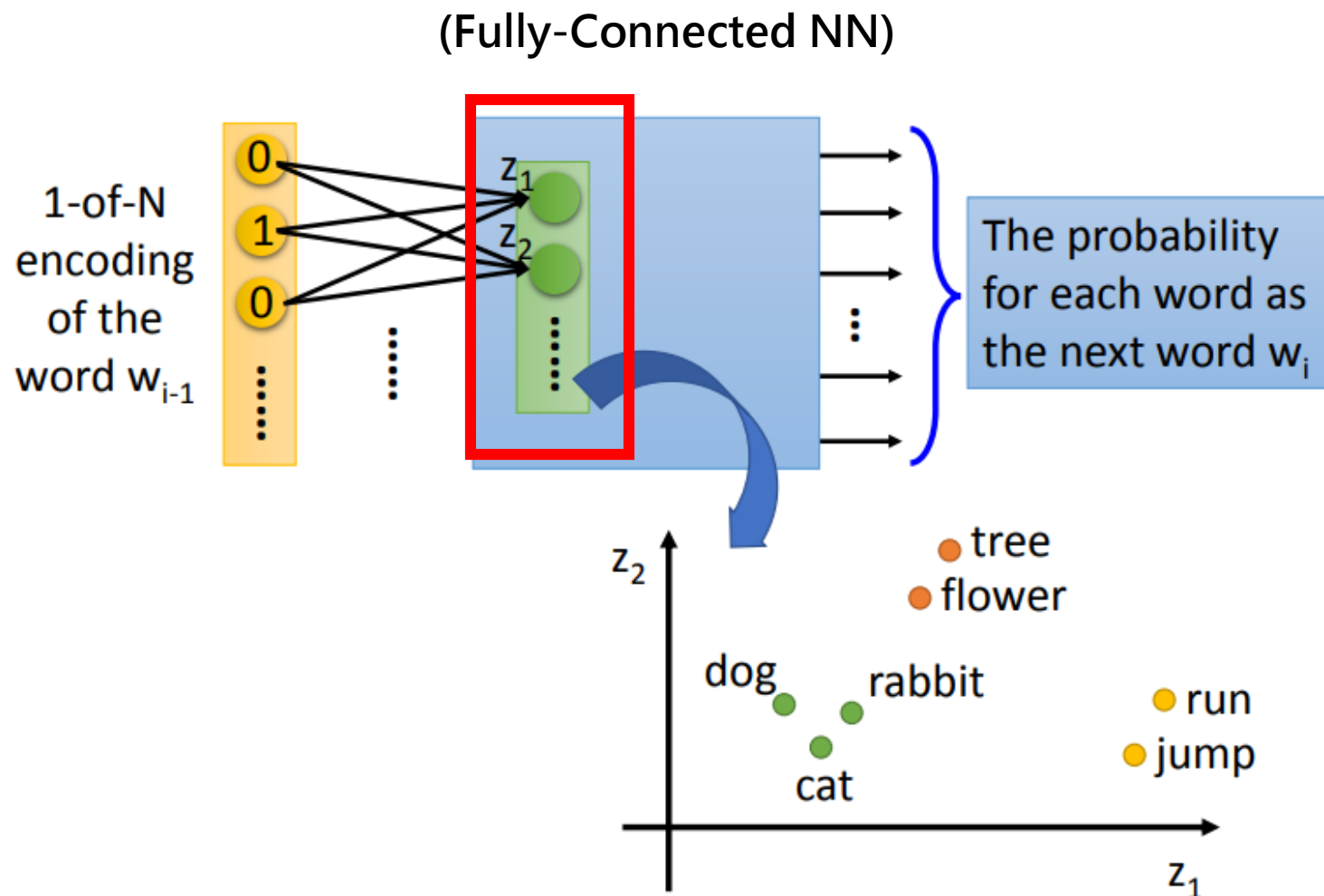
... .. W_{i-2} W_{i-1} W_i



降維:

可以從圖中發現，word embedding其實就是Hidden Layer第一層的值。

而假如我在這層Layer設計了N個Neurons，就可以將高維度的1-of-N Encoding降至N維。

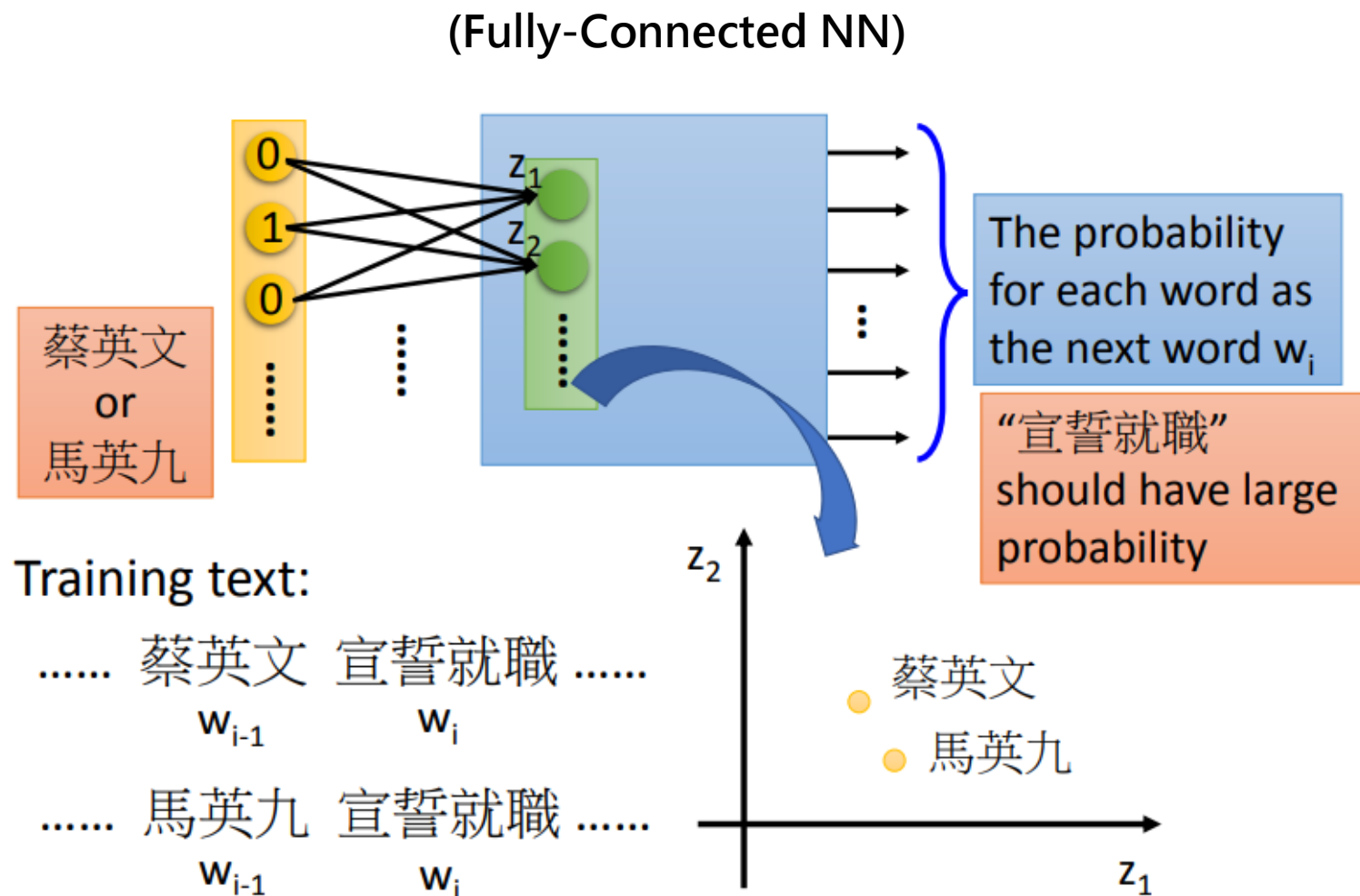


Word Embedding (Training)

透過上下文理解詞與詞的關聯:

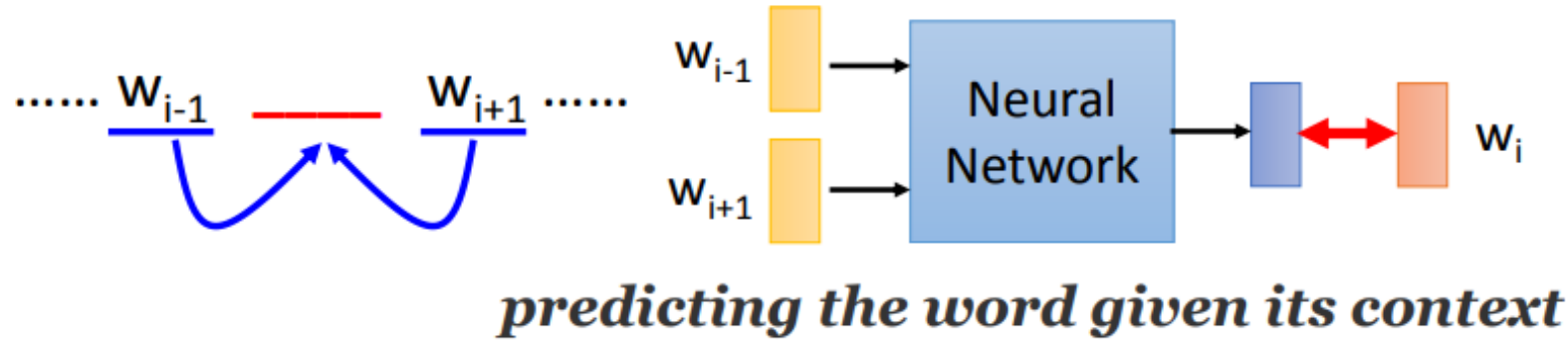
可以從這個例子看出，雖然input給machine的詞不一樣，但我們都希望它經過hidden layer後能夠output出“宣誓就職”。

我們希望machine學到的是，即使input不同，但是希望的產出一樣。因此在經過第一層hidden layer的weight transformation後，“蔡英文”、“馬英九”的word embedding會較為接近。

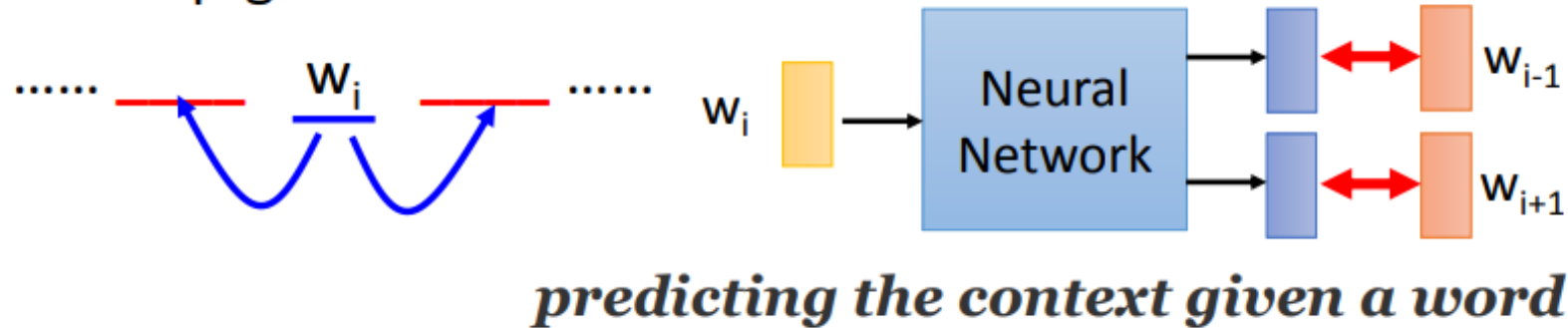


Word Embedding (Training)

- Continuous bag of word (CBOW) model



- Skip-gram



Word Embedding (缺點)

- 同樣的字詞，會因為上下文不同而有不同的意思，但word embedding無法體現

They stood on the river **bank**.

Deposit your money in the **bank**.

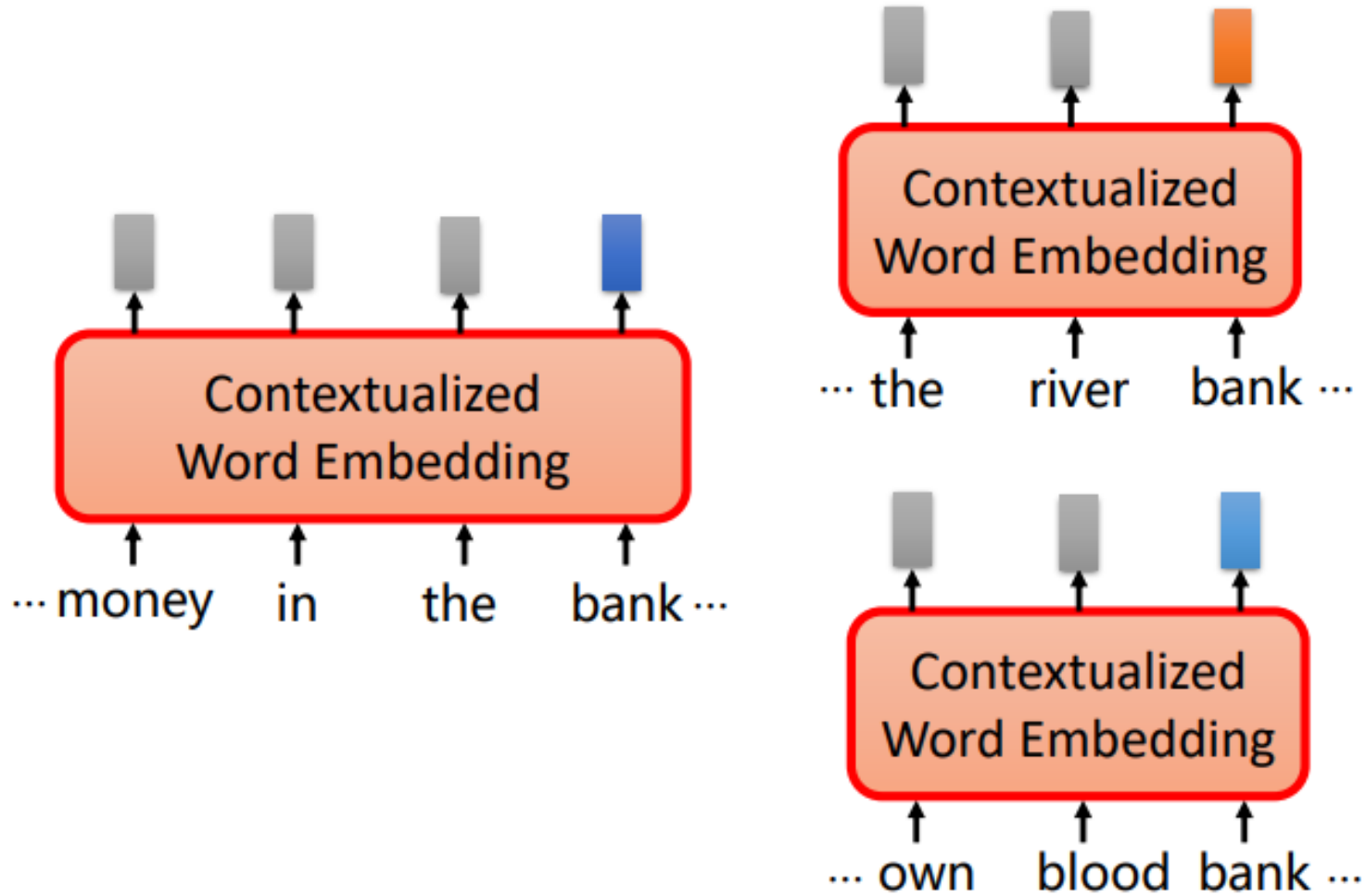
The hospital has its own blood **bank**.

- 上述這三個句子都存在著 “bank ” 這個詞彙，且隨著上下文的改變有著不同的意思。但如果給予這三個句子裡的 “bank ” 同樣的word embedding，會有偏誤



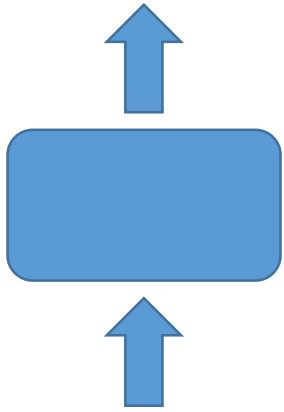
Contextualized Word Embedding

Contextualized Word Embedding



Self-attention

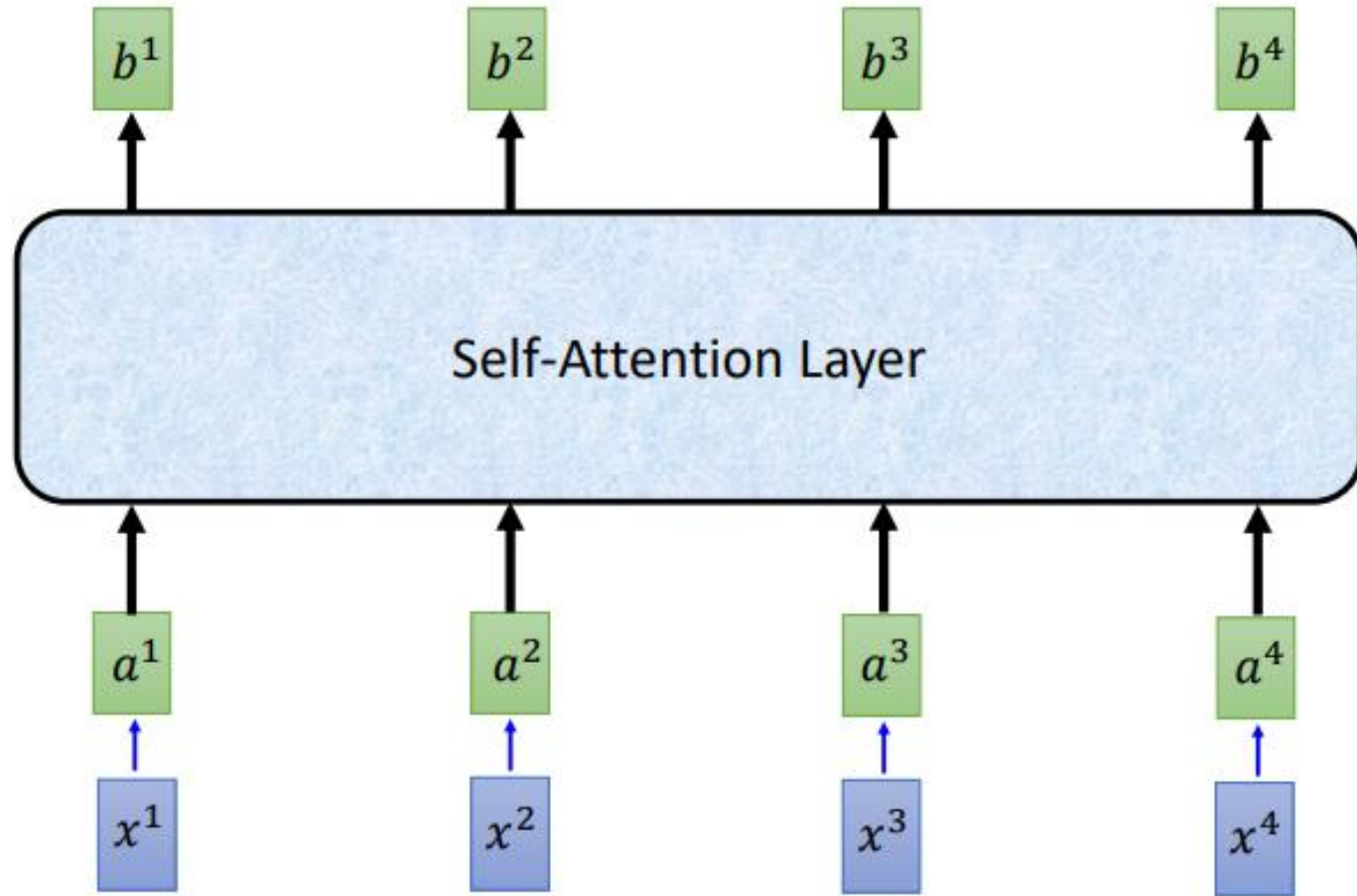
Output:
Contextualized
Word Embedding



Word Embedding:
 $a^i = Wx^i$



Input:
1-of-N Encoding



Self-attention

q (query): to match others

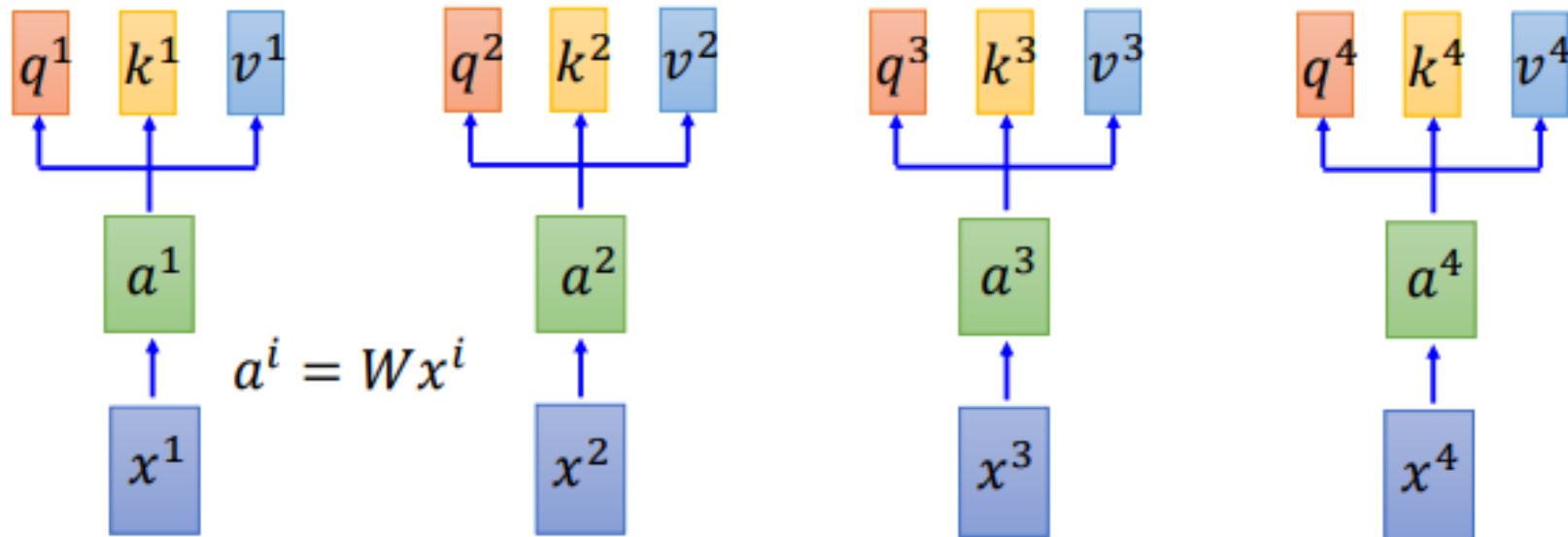
$$q^i = W^q a^i$$

k (key): to be matched

$$k^i = W^k a^i$$

v: information to be extracted

$$v^i = W^v a^i$$



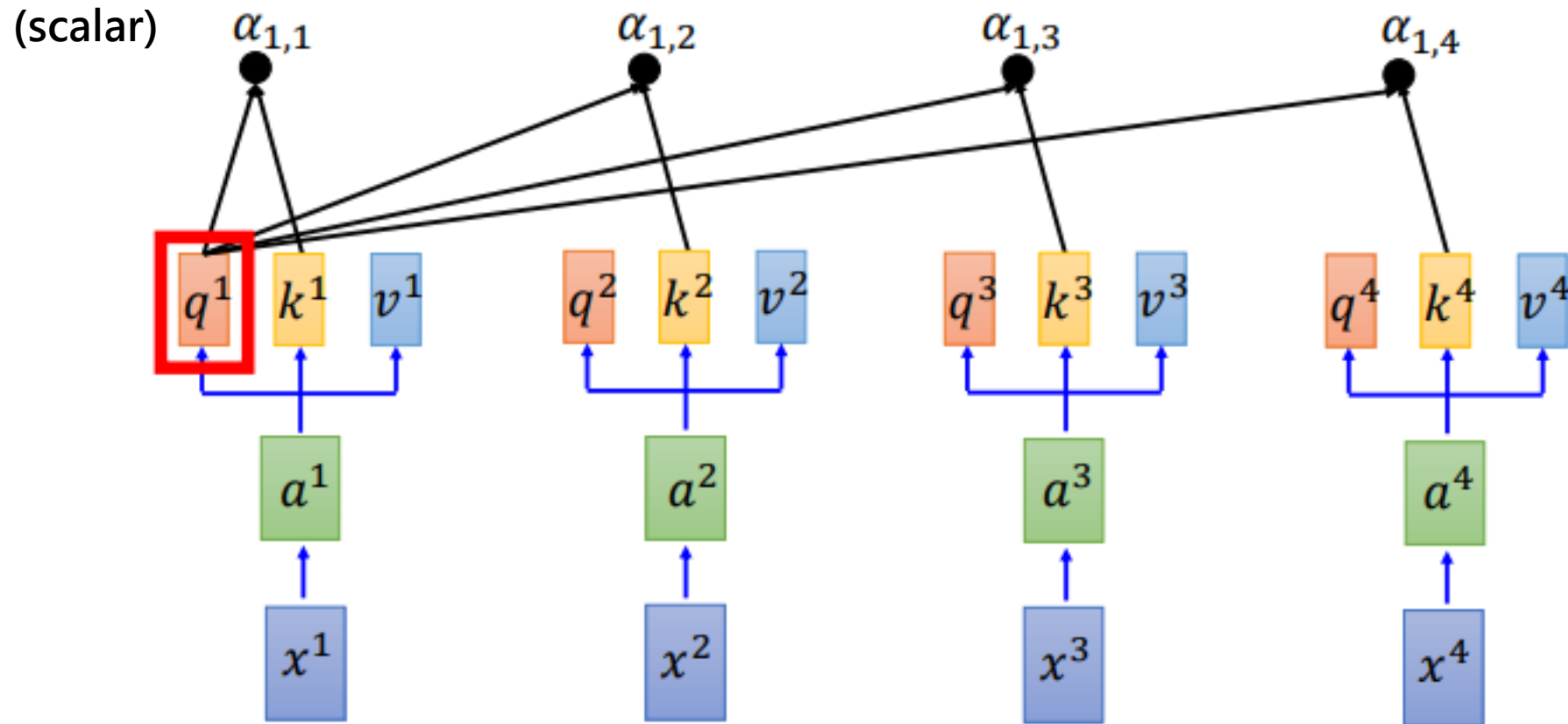
Self-attention

- 拿每一個 query 去對每一個 key 做 attention (算匹配程度 / Similarity)

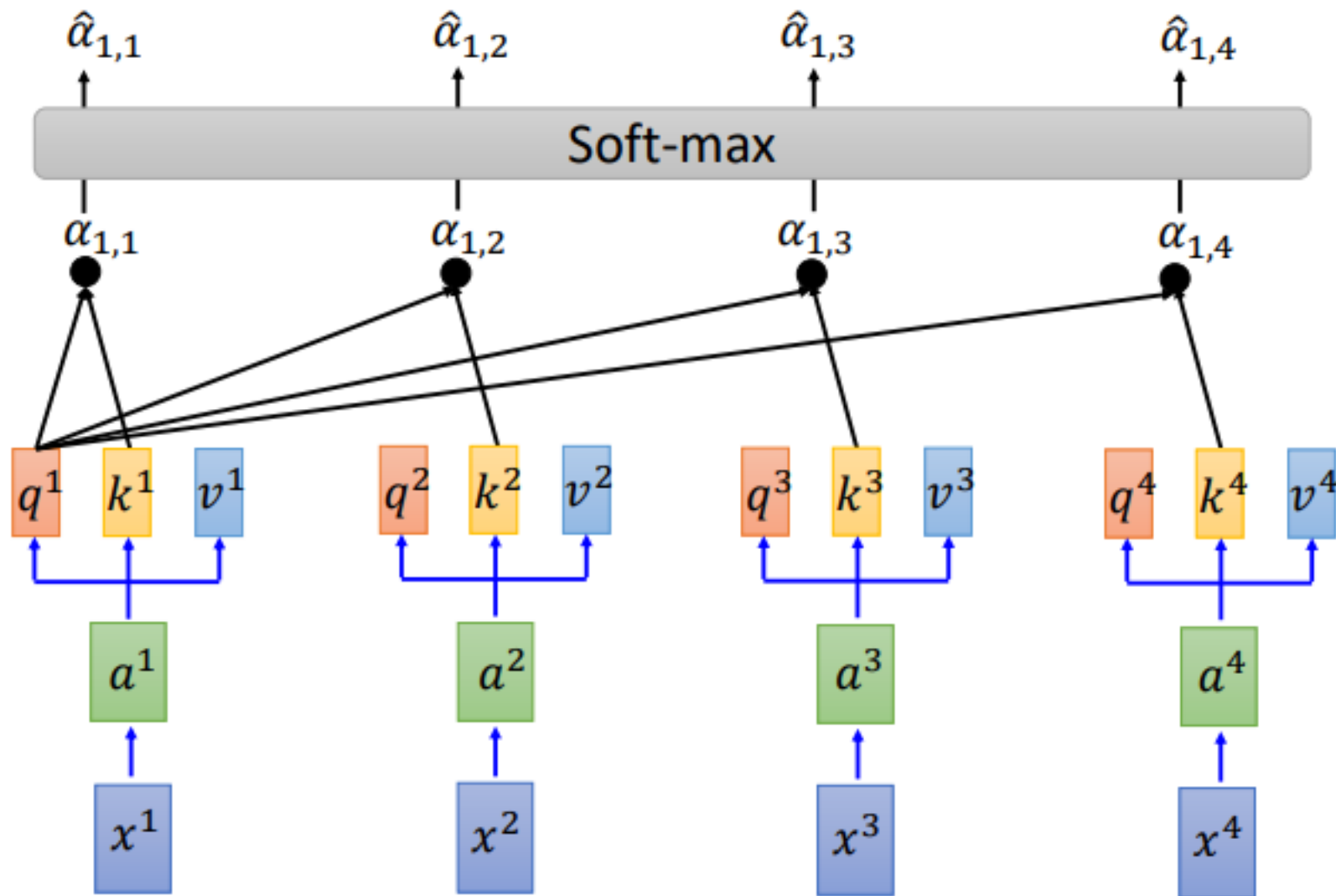
Note: cosine similarity

$$\cos(\theta) = \frac{q \cdot k}{\|q\| \|k\|}$$

Scaled Dot-Product Attention: $\alpha_{1,i} = q^1 \cdot k^i / \sqrt{d}$ (d is the dim of q and k)



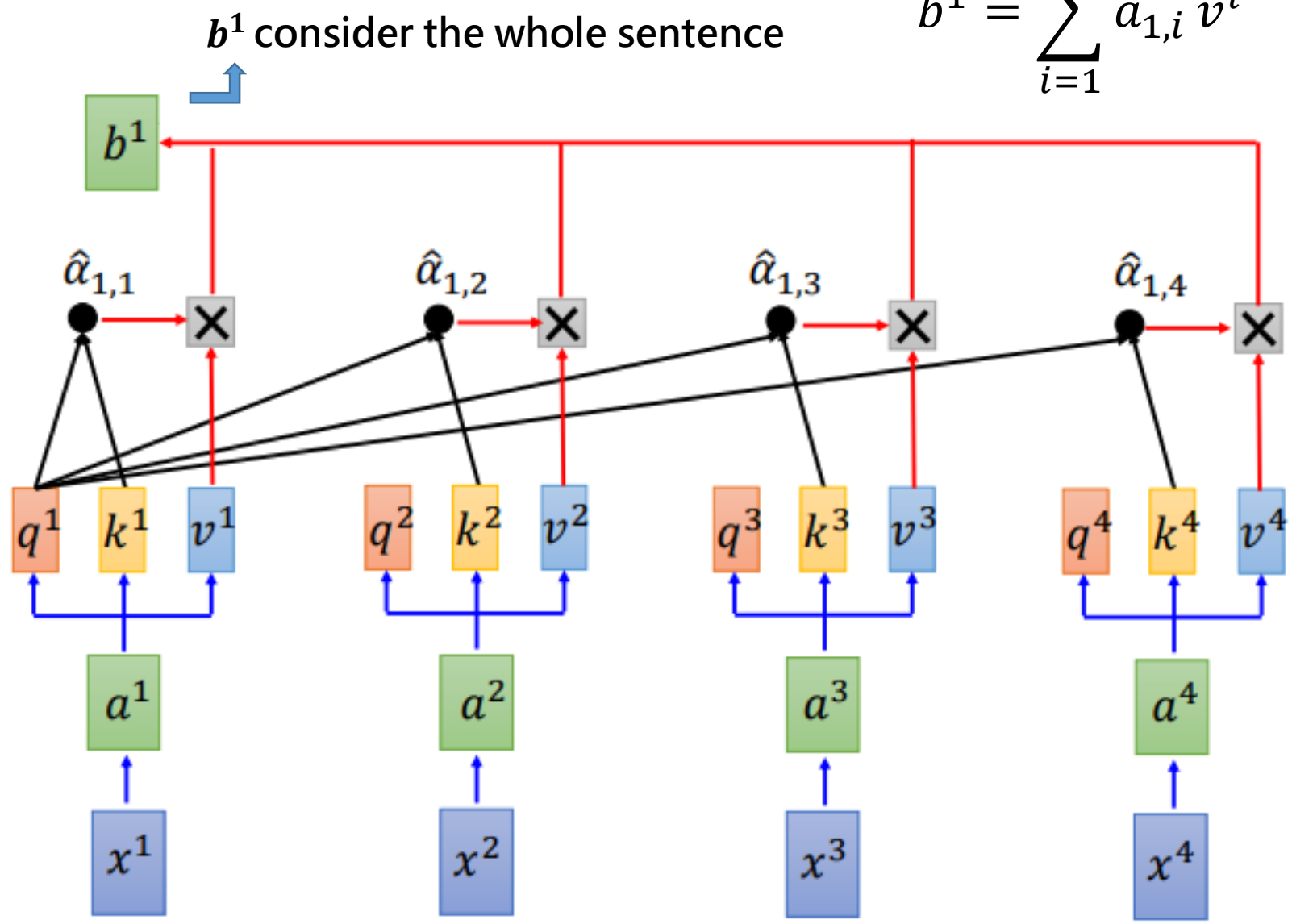
Self-attention



使 $\hat{a}_{1,i}$ 皆大於0且和為1
weighted sum的概念

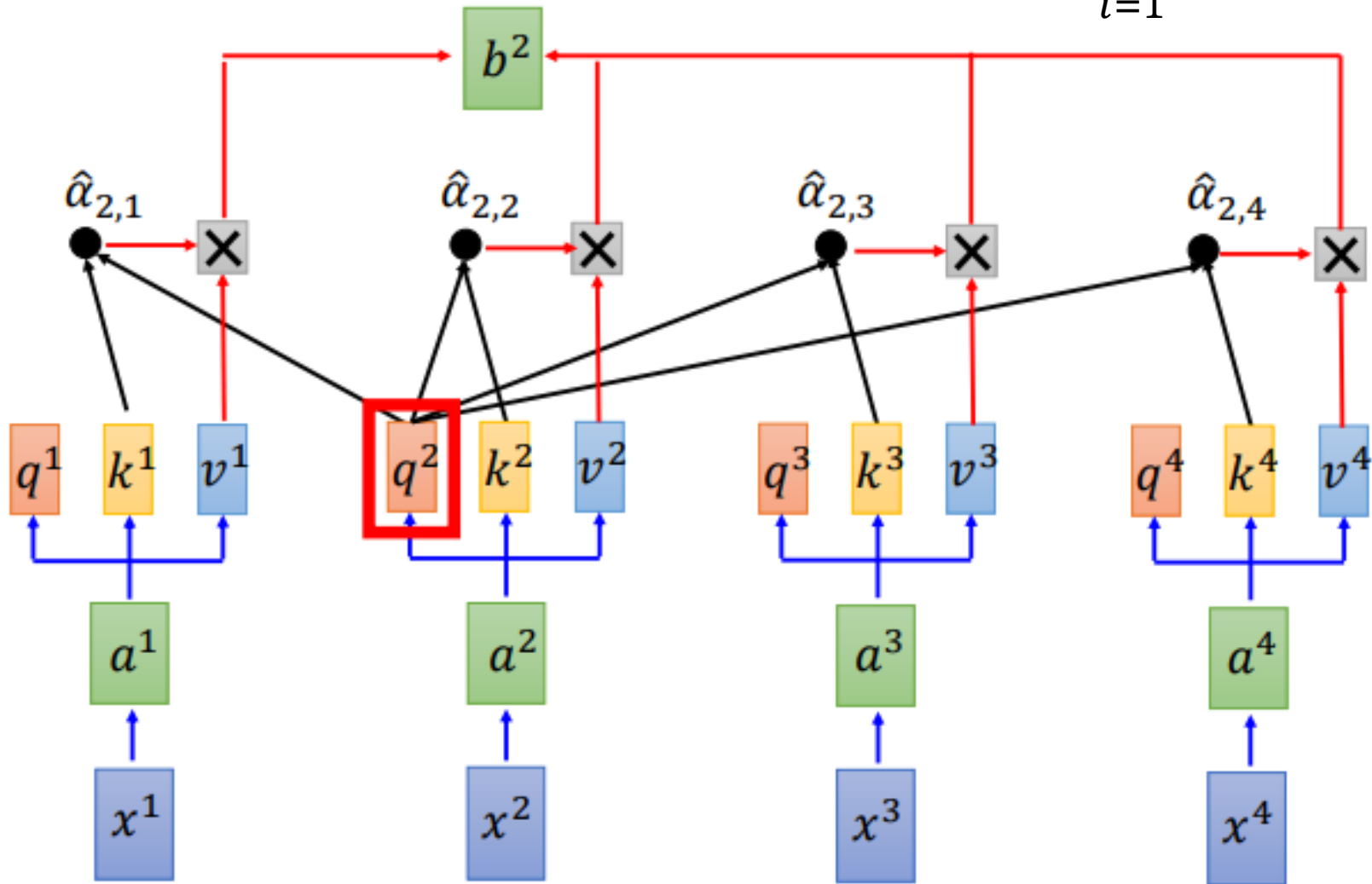
Self-attention

$$b^1 = \sum_{i=1}^4 \hat{a}_{1,i} v^i$$



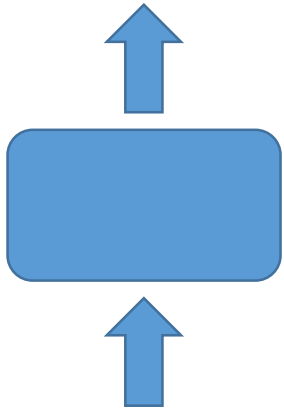
Self-attention

$$b^2 = \sum_{i=1}^4 \hat{a}_{2,i} v^i$$



Self-attention

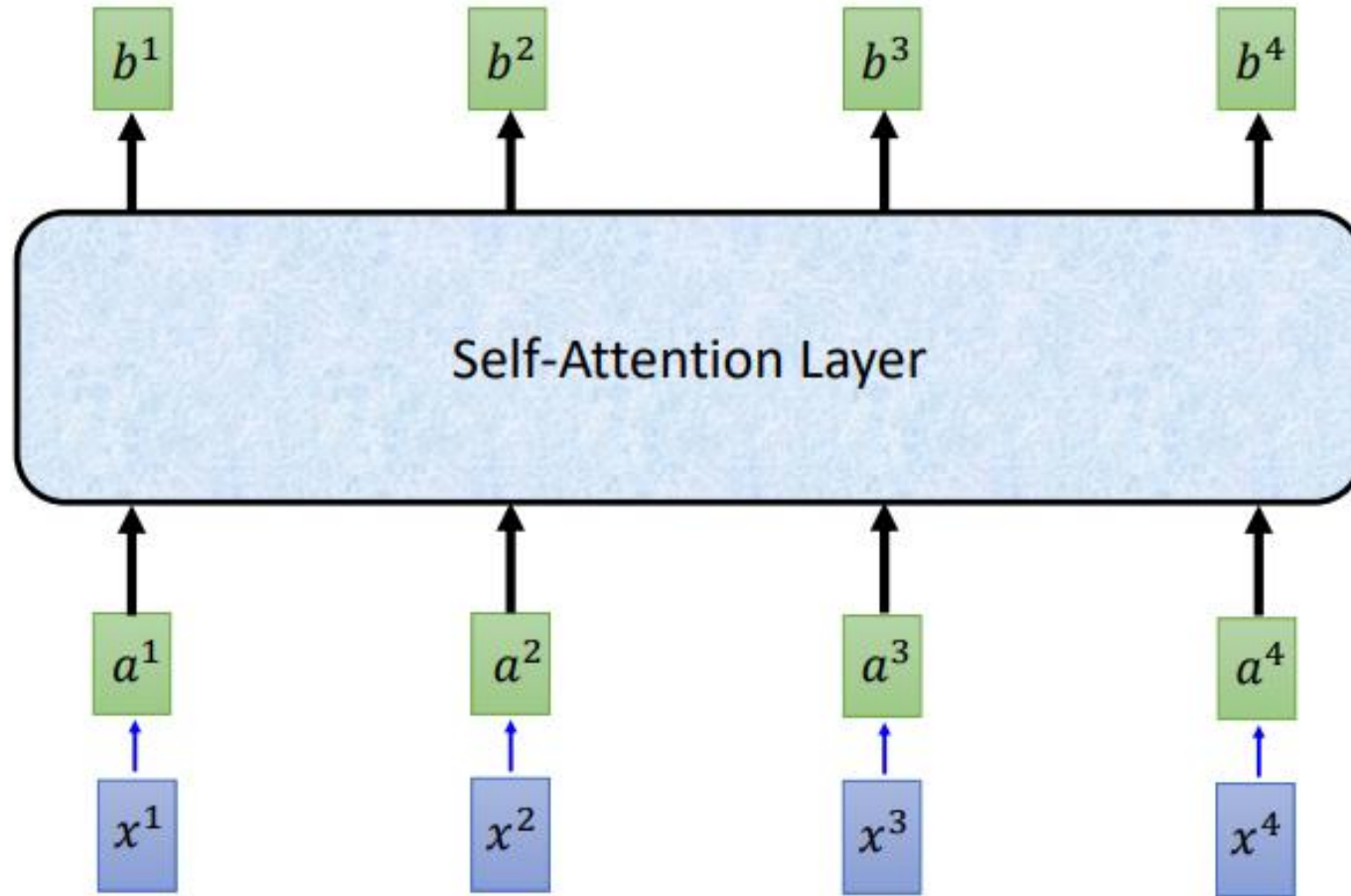
Output:
Contextualized
Word Embedding



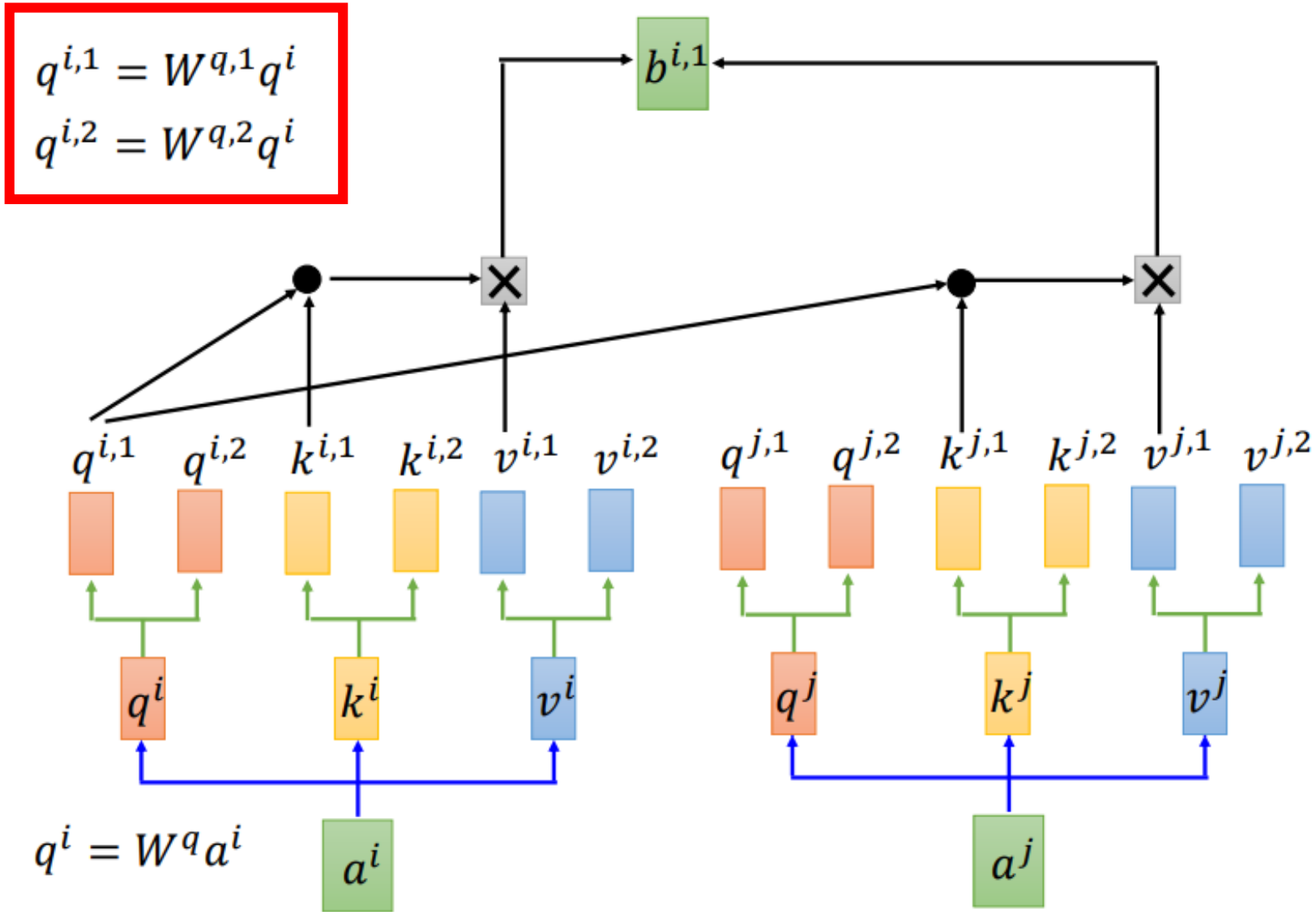
Word Embedding:
 $a^i = Wx^i$



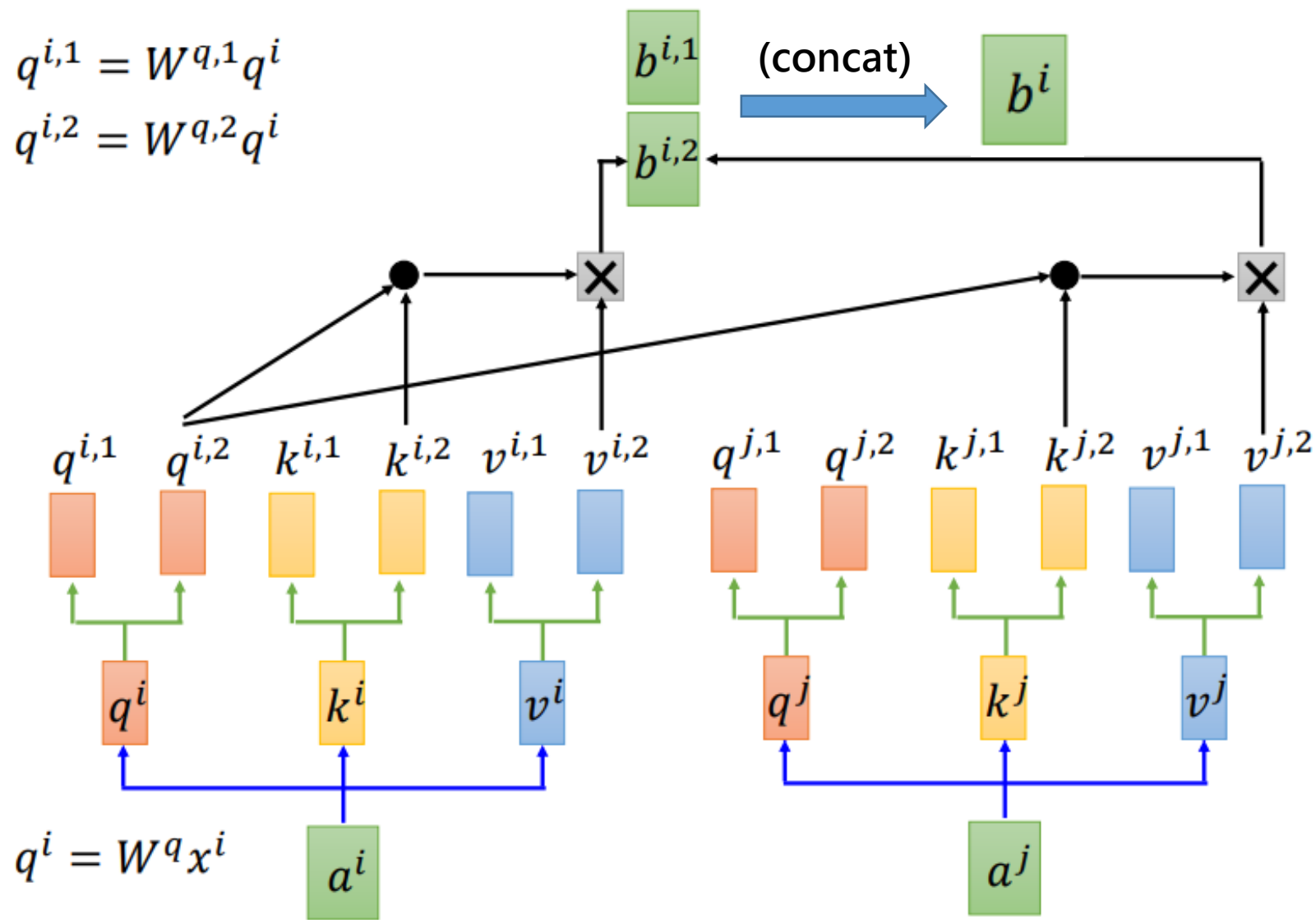
Input:
1-of-N Encoding



Multi-head Self-attention (2 heads for example)



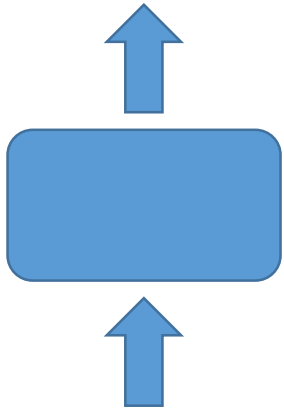
Multi-head Self-attention (2 heads for example)



- 實作上 heads 的數量可以用 CV 去調整
- 設定多個 heads 的好處: 不同的 head 可以擷取到不同段落的語意。舉例來說，第一個 head 可能是擷取到句子前半段的語意；第二個 head 可能是擷取到句子後半段的語意。

Self-attention

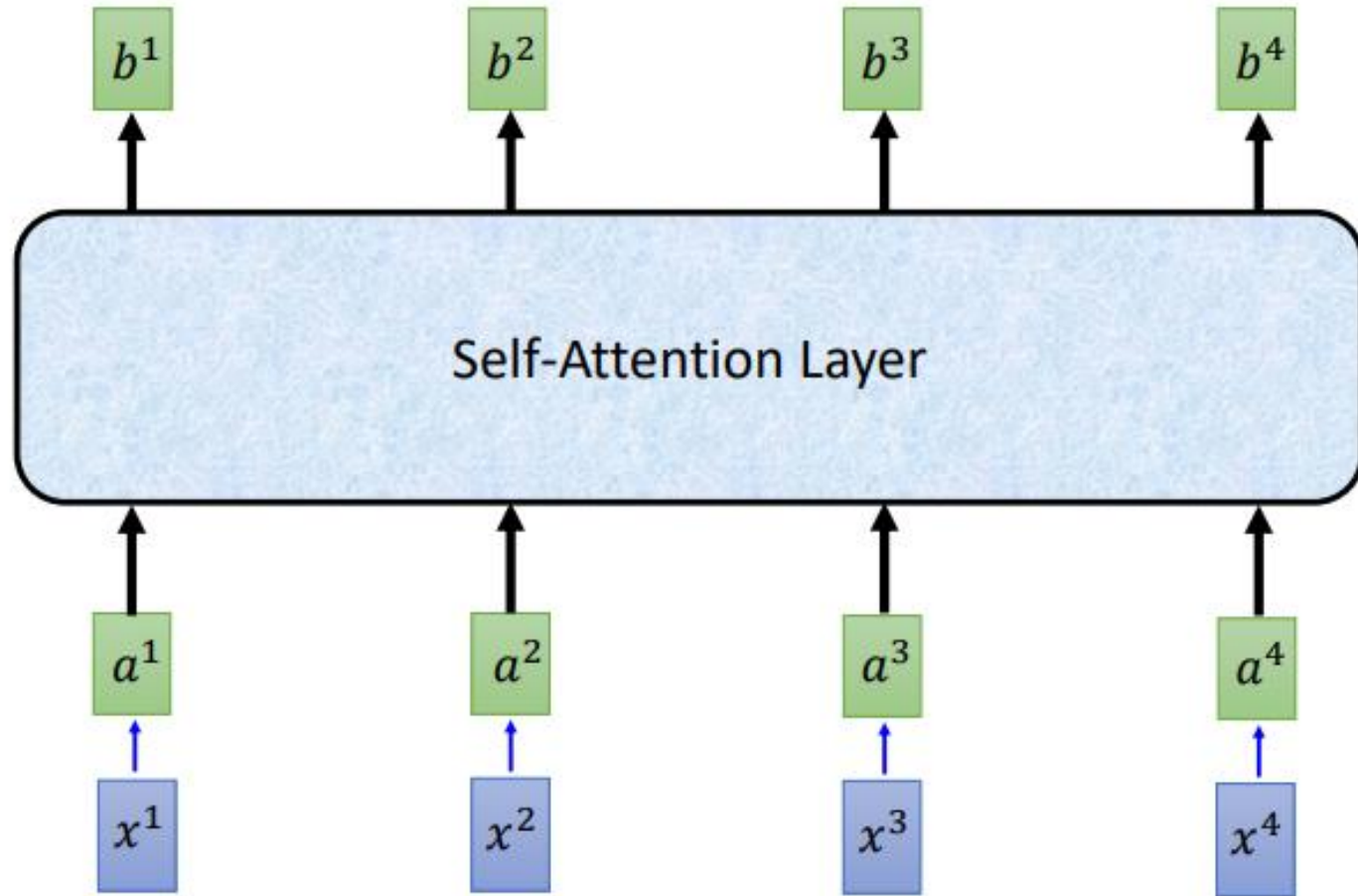
Output:
Contextualized
Word Embedding



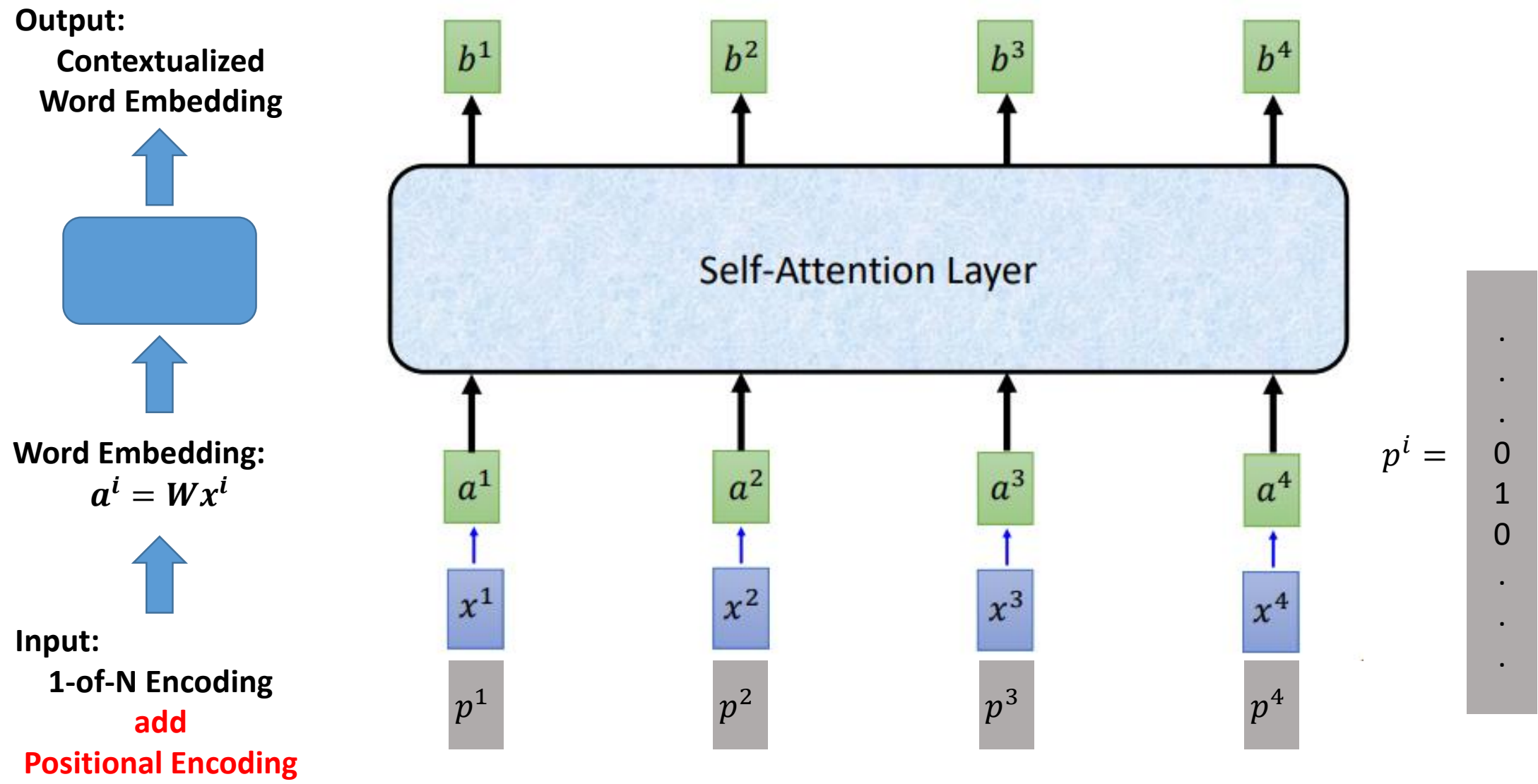
Word Embedding:
 $a^i = Wx^i$



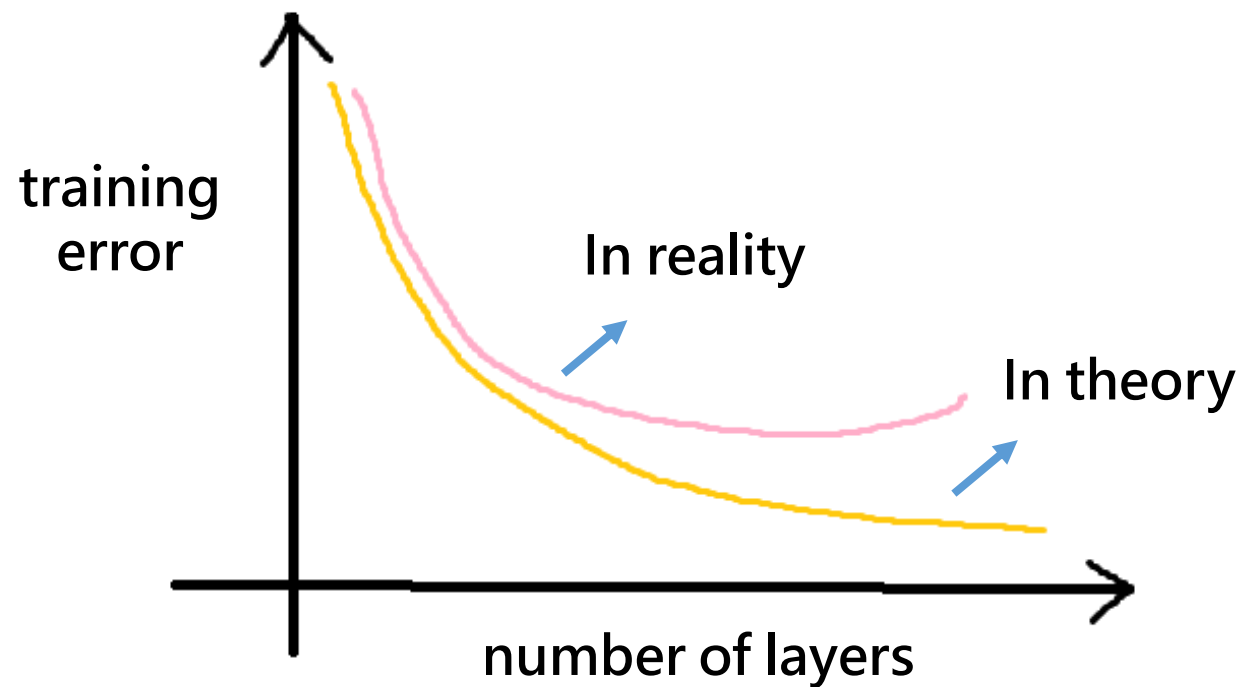
Input:
1-of-N Encoding



Self-attention



Residual Network

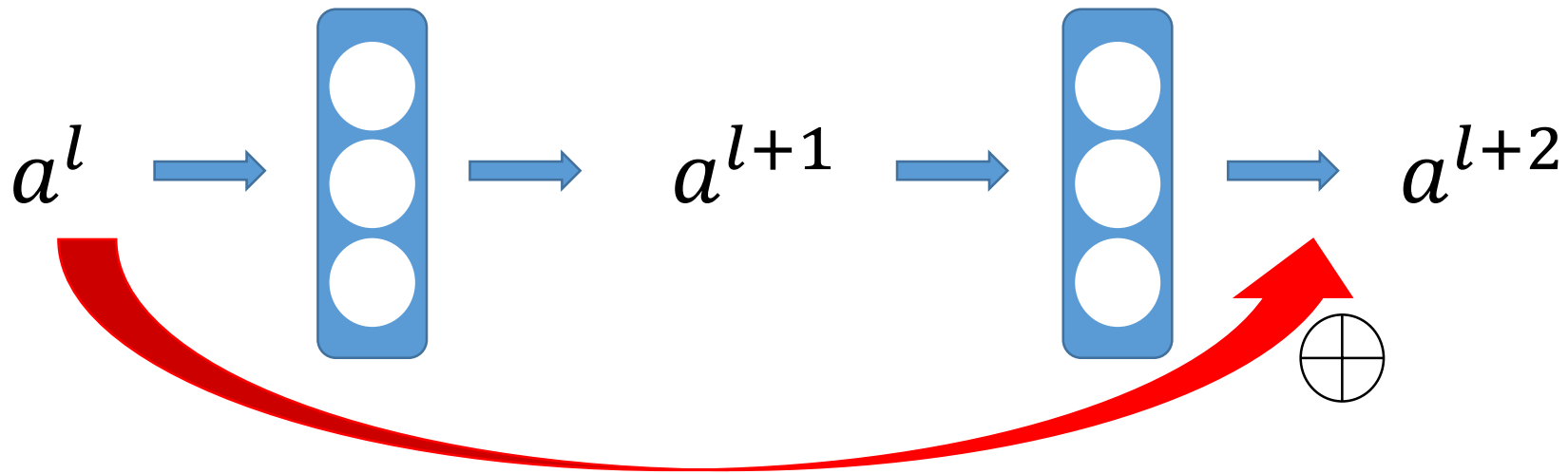


- 我們在建構模型架構時，一般不會知道要建的多少層。因此我們需要尋找一個辦法，讓我們的模型在建構更多層 layer 時，至少在訓練集的表現不會變的更差。

- Short Cut / Skip Connection

Residual Network

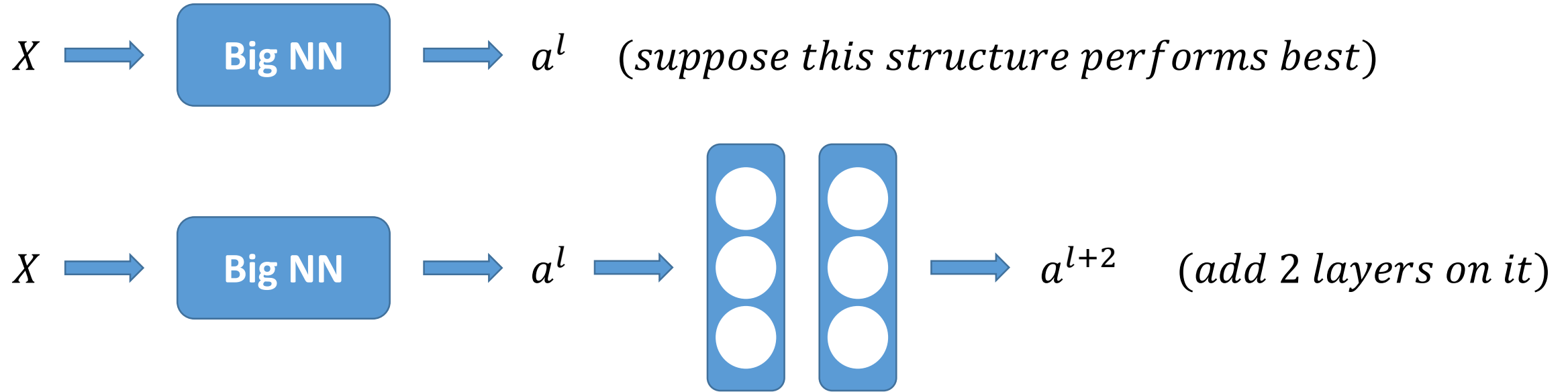
➤ Short Cut / Skip Connection



$$Z^{l+1} = W^{l+1} \cdot a^l + b^{l+1} \quad , \quad a^{l+1} = g(Z^{l+1})$$

$$Z^{l+2} = W^{l+2} \cdot a^{l+1} + b^{l+2} \quad , \quad a^{l+2} = g(Z^{l+2} + a^l)$$

Residual Network



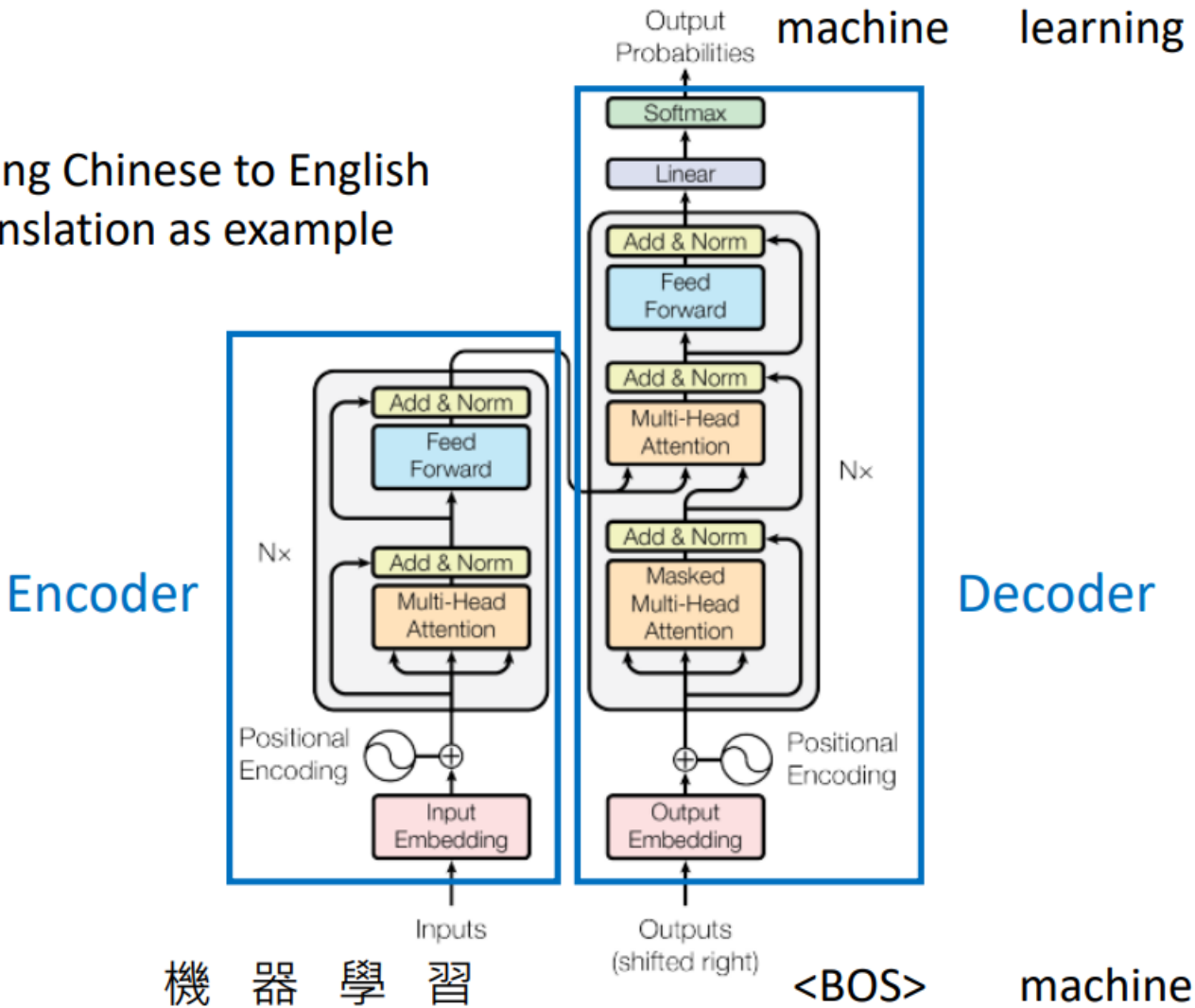
- 我們希望 machine 學到: 不應該加上這 2 層 layer (activation: ReLU)

$$a^{l+2} = g(Z^{l+2} + a^l) = g(W^{l+2} \cdot a^{l+1} + b^{l+2} + a^l)$$

- W^{l+2} 、 b^{l+2} 是模型的權重，是需要 learn 出來的。我們加上 short cut 後，
假如模型將這兩組權重 learn 為 0，則 $a^{l+2} = a^l$

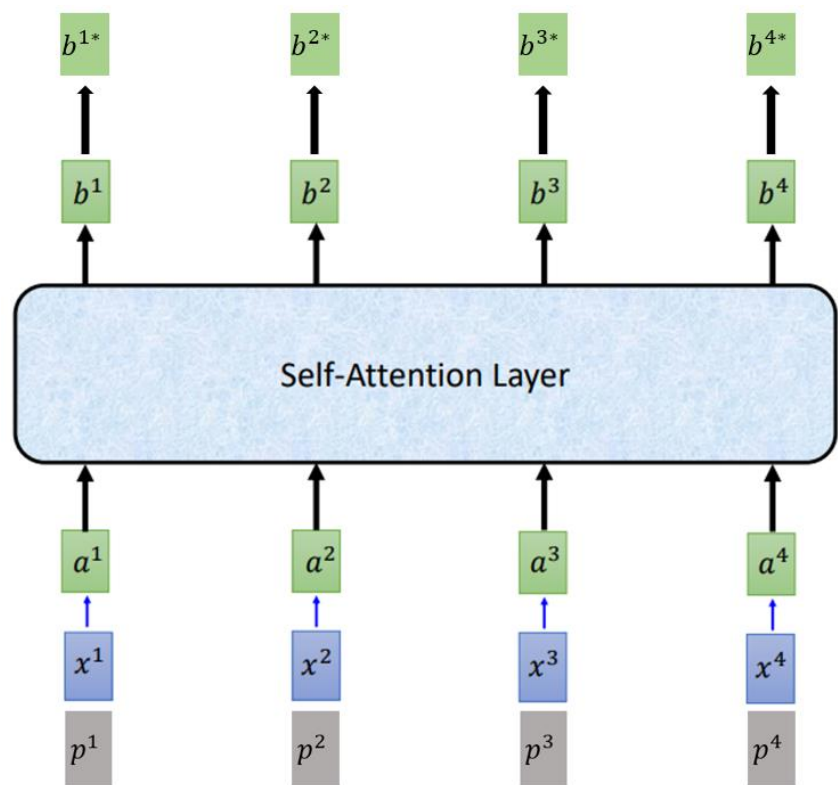
Transformer

Using Chinese to English translation as example

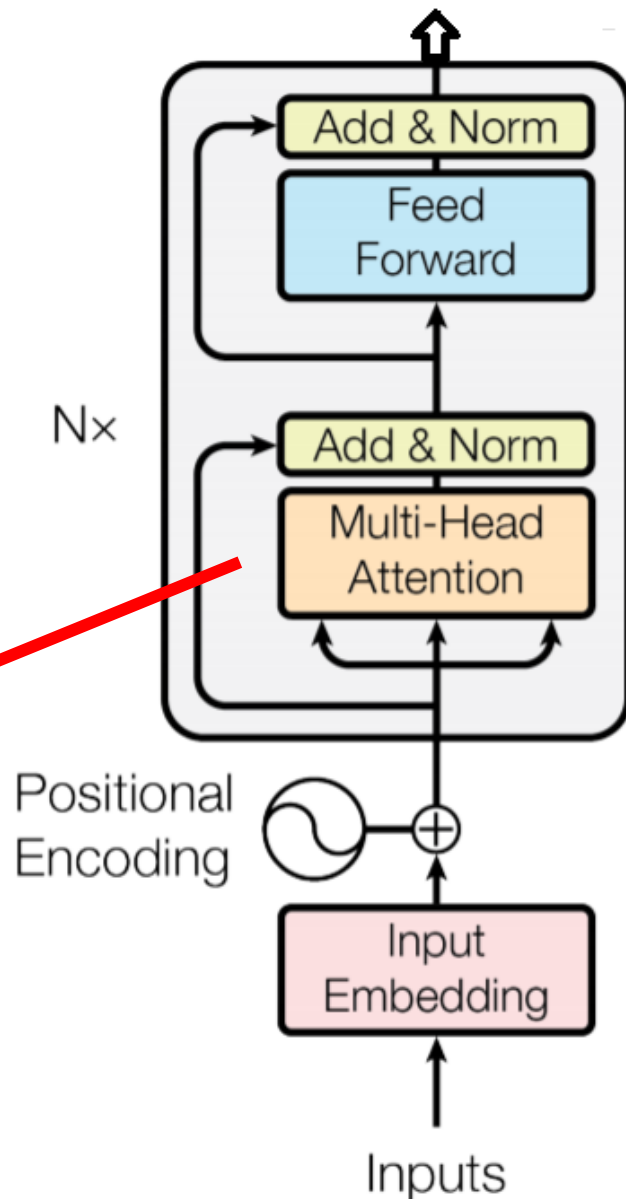


BERT 就是多層Transformer 的 Encoder

- Add 的用意和前面提到的 Residual Network 一樣，因為模型會建的很深，我們希望 machine 自己去學習哪些 layer 是不需要用到的。

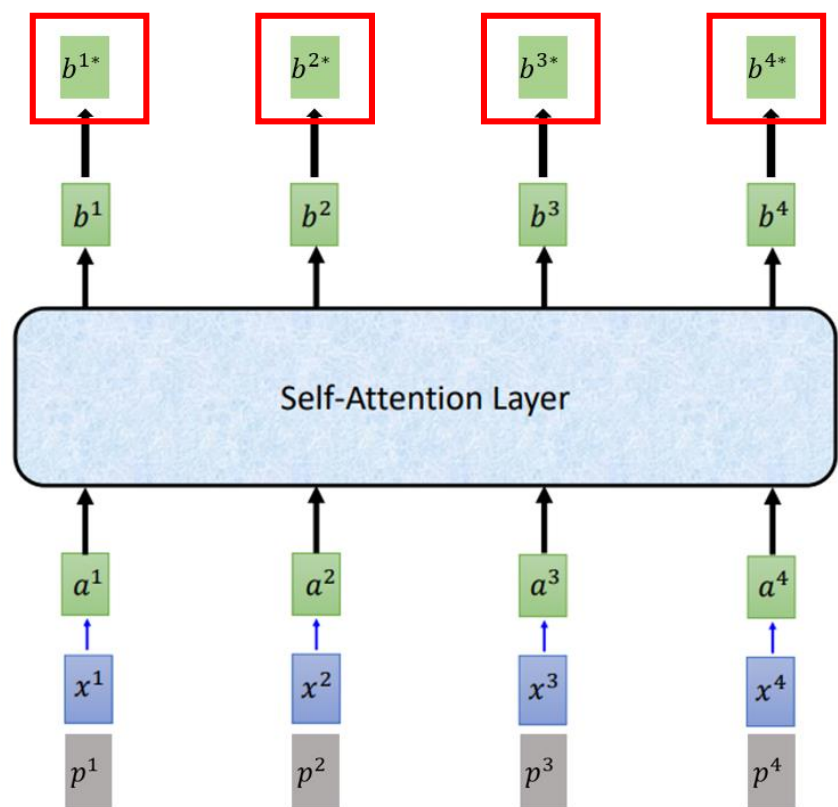


➤ Add
$$b^{i*} = a^i + b^i$$

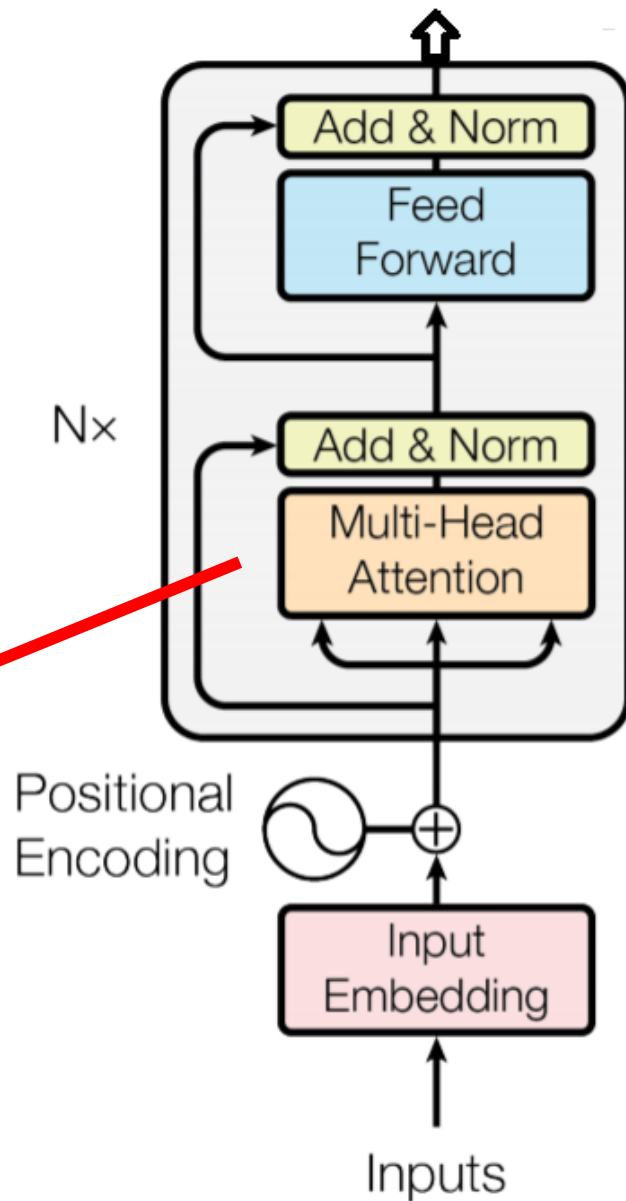


BERT 就是多層Transformer 的 Encoder

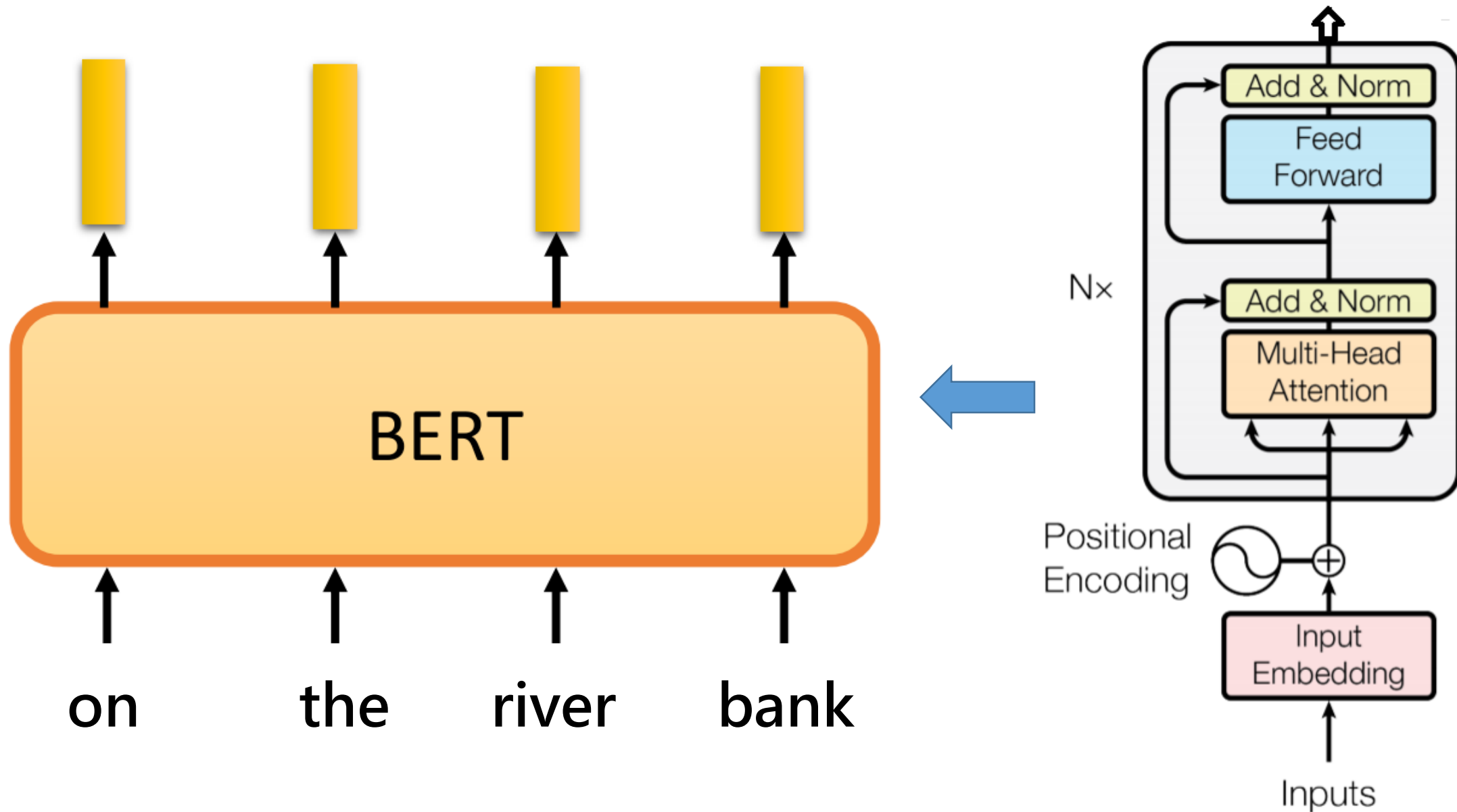
- 因為每一層 layer 的 output 就是下一層 layer 的 input。將每一層 layer 的 output 標準化，可以使我們在做 gradient descent 時，不會受權重的初始值影響太大。



➤ Norm: 將每一個 b^{i*} 裡的 elements 標準化至 $\mu = 0, \sigma = 1$

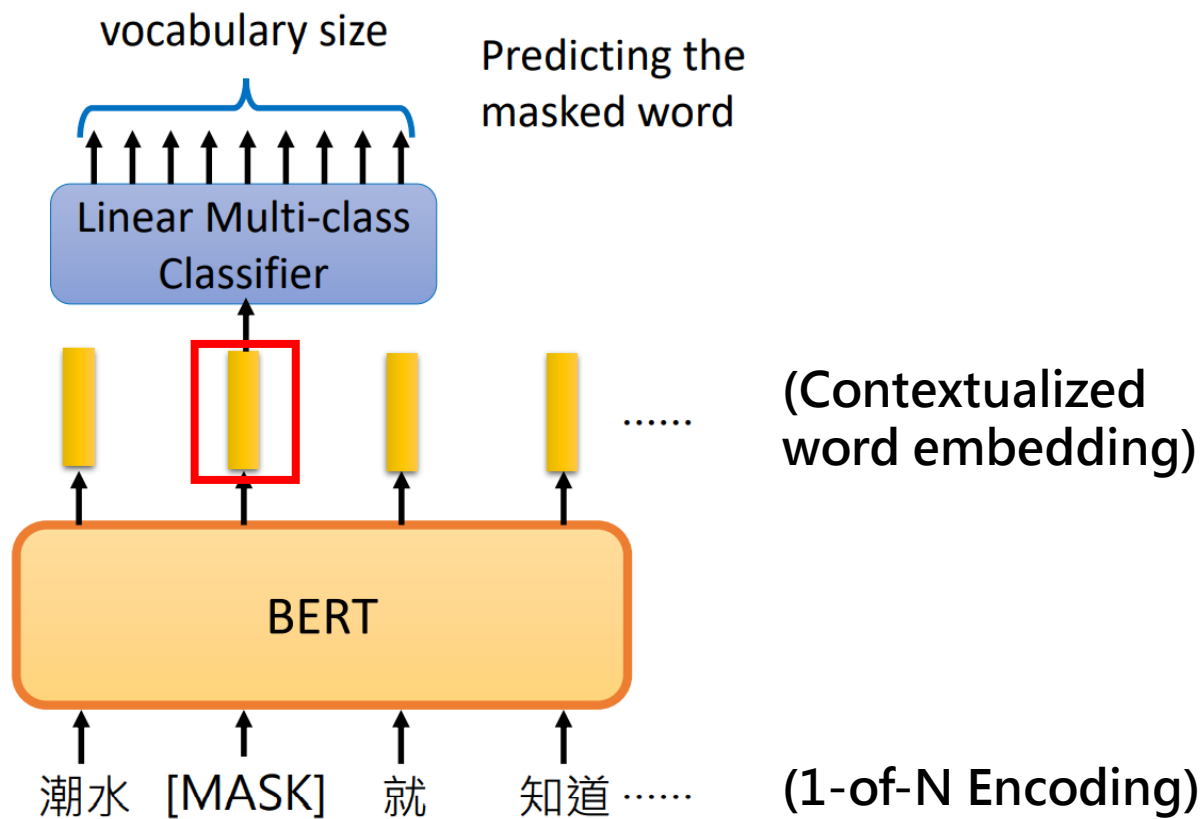


BERT 就是多層Transformer 的 Encoder



Training of BERT (1)

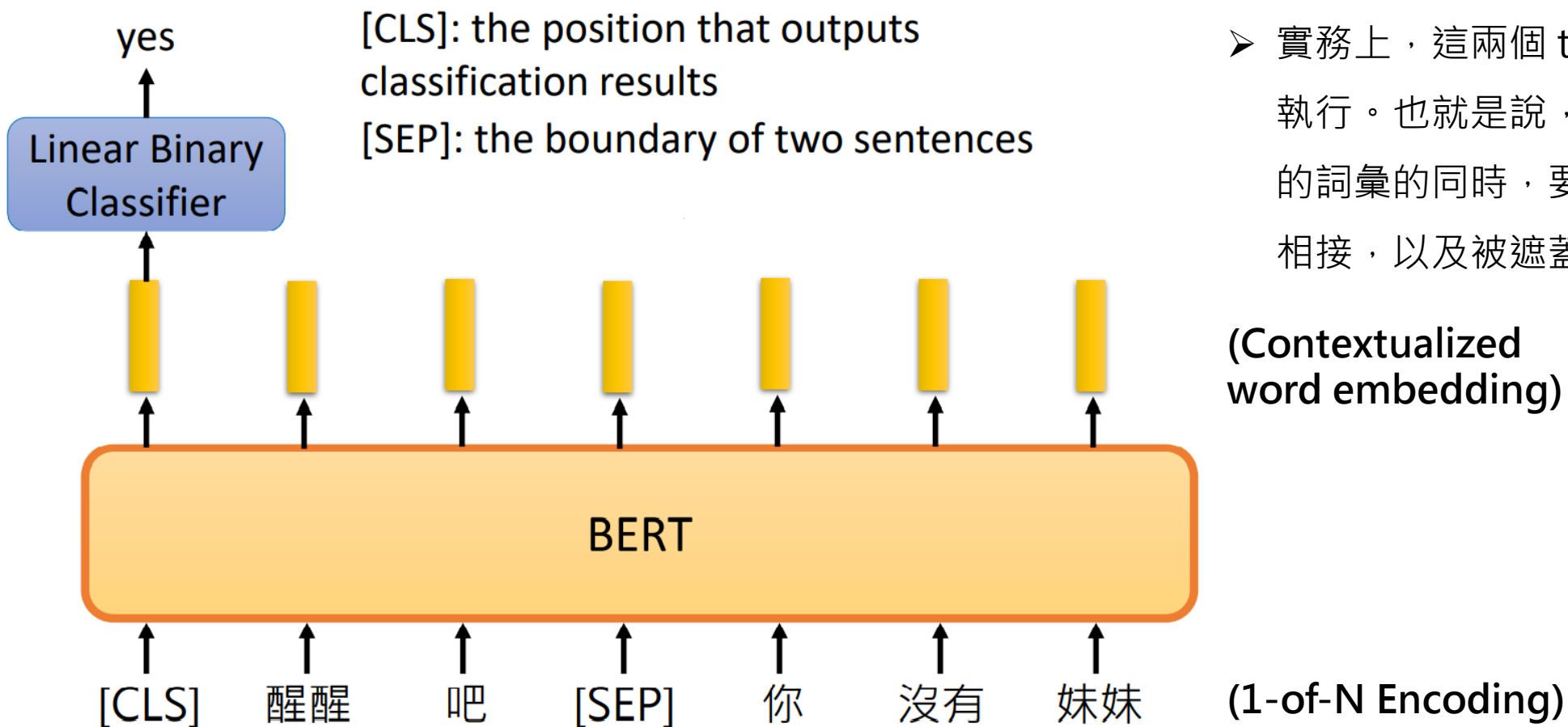
- **克漏字填空**: 將每一筆 training set 中的句子，隨機蓋掉 15% 的字，將蓋掉的地方改為一個 special token: **MASK**。讓 machine 去預測這個被蓋掉的字是什麼。



- 因為克漏字這個 task 是將 MASK 這個特殊的 token 丟入 BERT 後產生的 embedding，再丟入一個相對較為簡單的線性分類器裡面 (logistic regression)。因此這個 word embedding 必須要帶有足夠的資訊，才能使機器正確猜出被蓋掉的詞彙。
- 另外在這個例子上，填入「退了」或「落了」其實都還算符合語意。因此這兩個詞彙就會有著較為相似的 embedding。

Training of BERT (2)

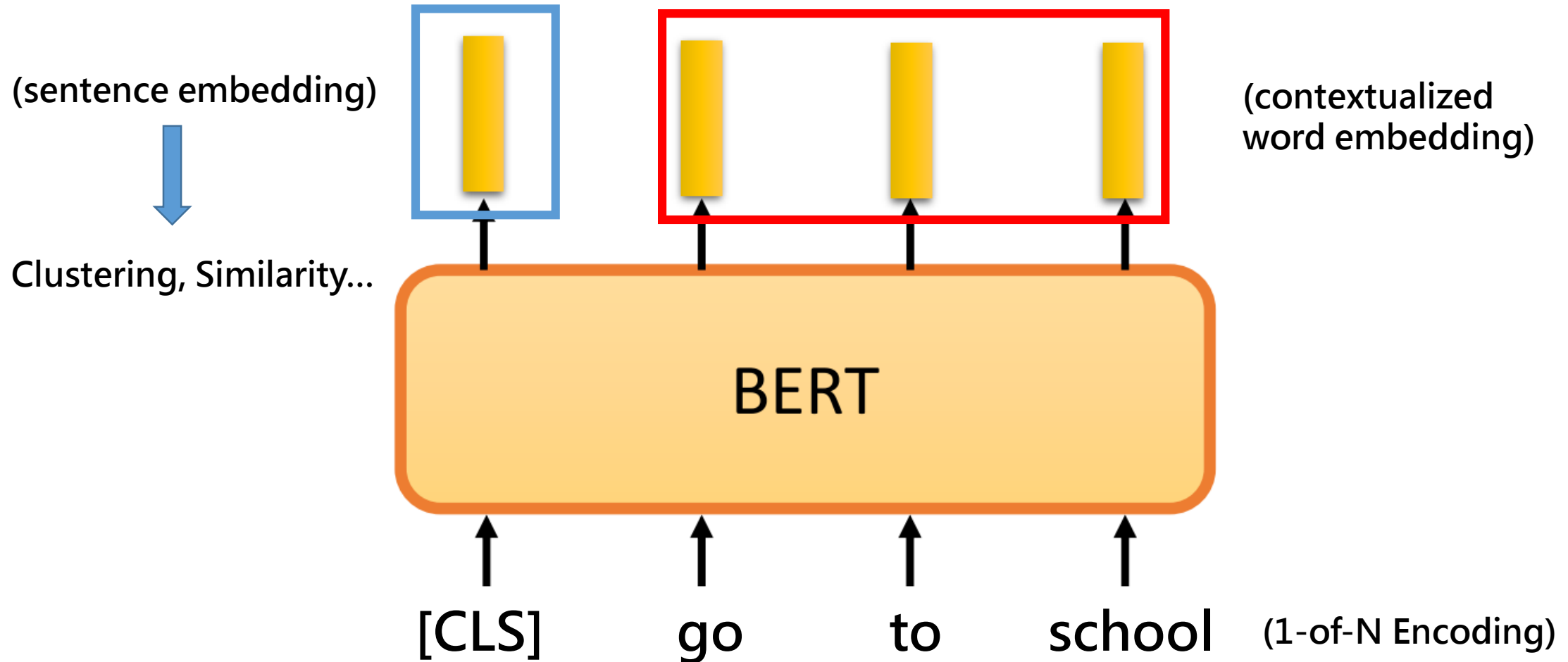
- 判斷第 2 個句子在原始文本中是否跟第 1 個句子相接



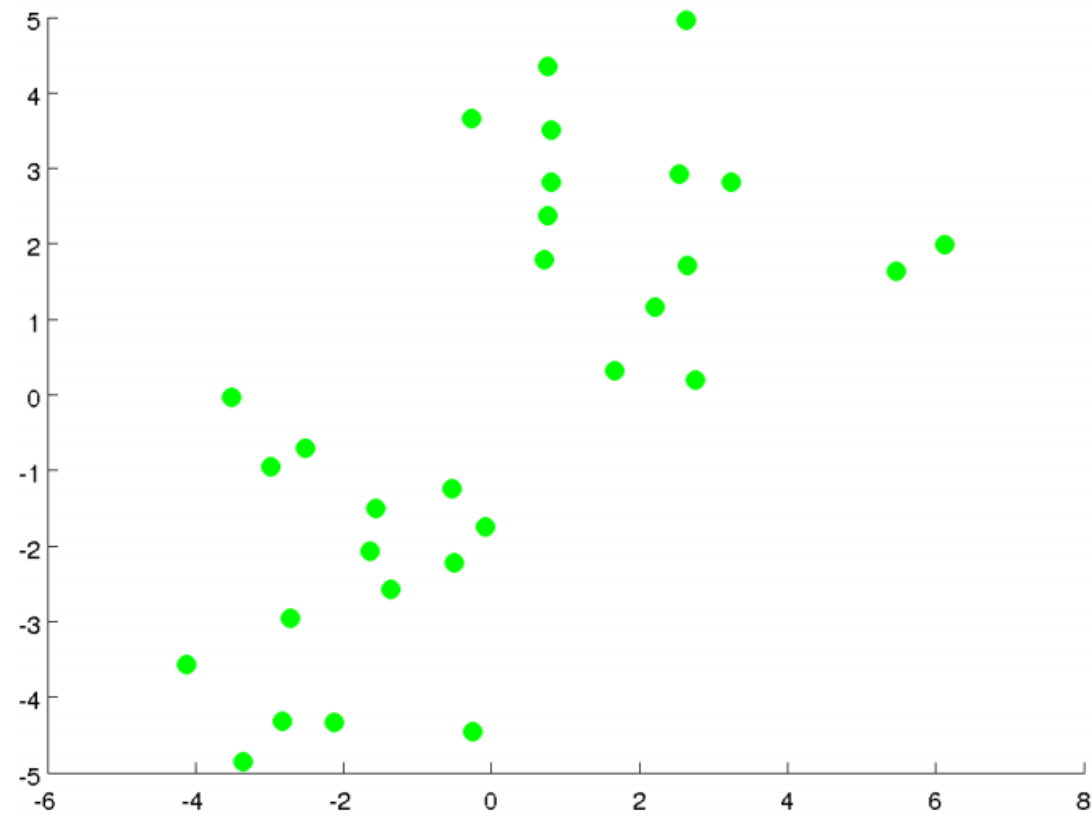
- 實務上，這兩個 training 的方法會一起執行。也就是說，BERT 會在被蓋掉 15% 的詞彙的同時，要去預測兩兩句子是否相接，以及被遮蓋掉的詞彙為何。

Usage of Pre-trained BERT

- sentence embedding & contextualized word embedding

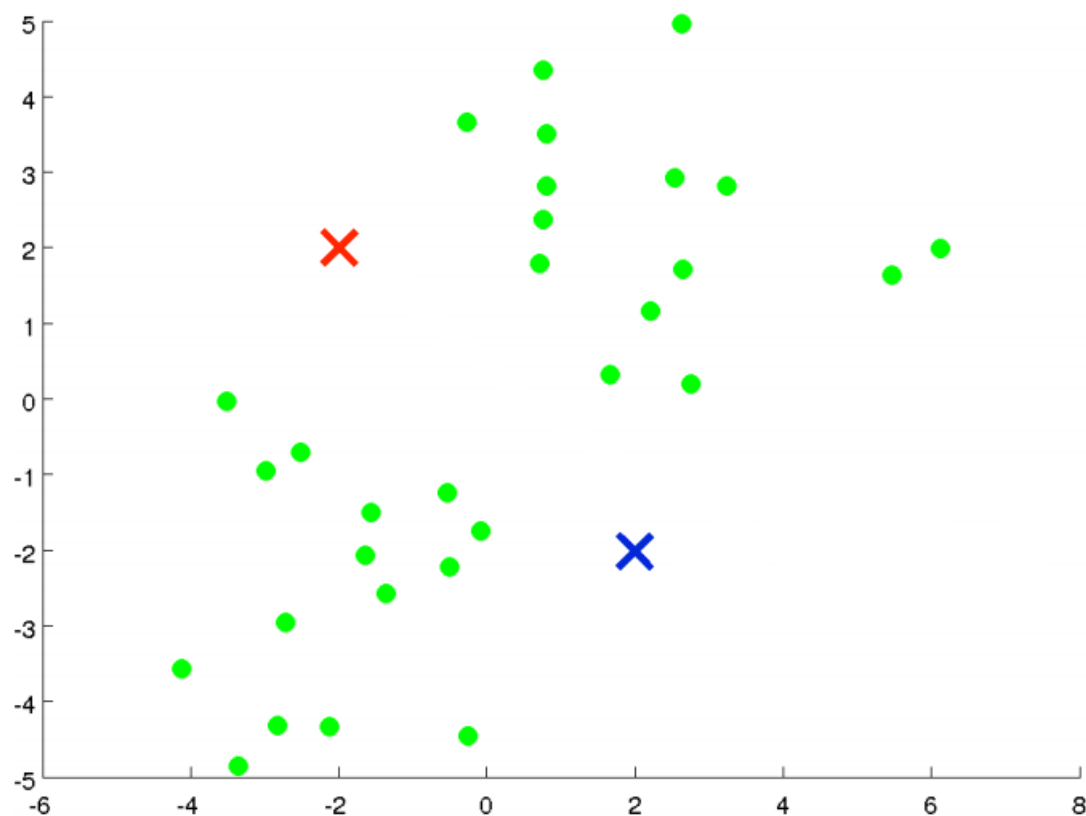


Usage of Pre-trained BERT: K-means



Usage of Pre-trained BERT: K-means

- 假設我們今天想要將資料分成兩群，則會先隨機產生兩個 cluster centroid
- 要分幾群依自己主觀選擇，也可以看分群效果如何去決定

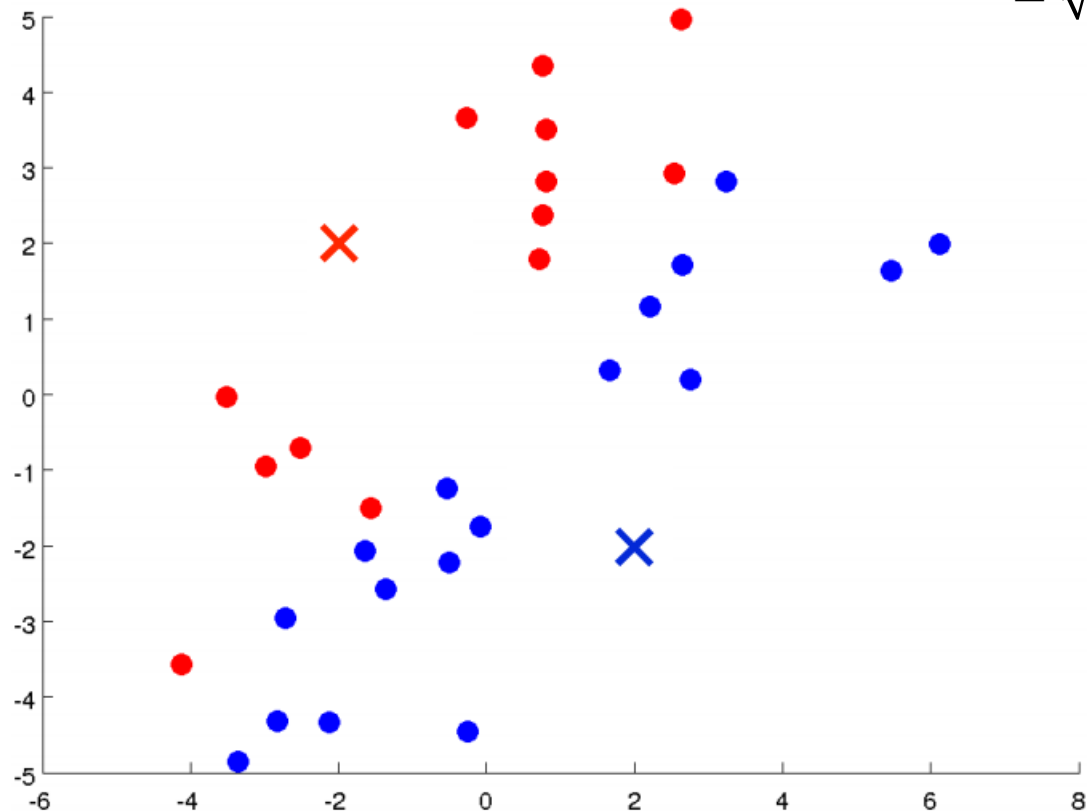


Usage of Pre-trained BERT: K-means

- 接著去計算每一個資料點和兩個 cluster centroid 的歐氏距離
- 將和某一 centroid 距離較近的資料點，歸類為同一群

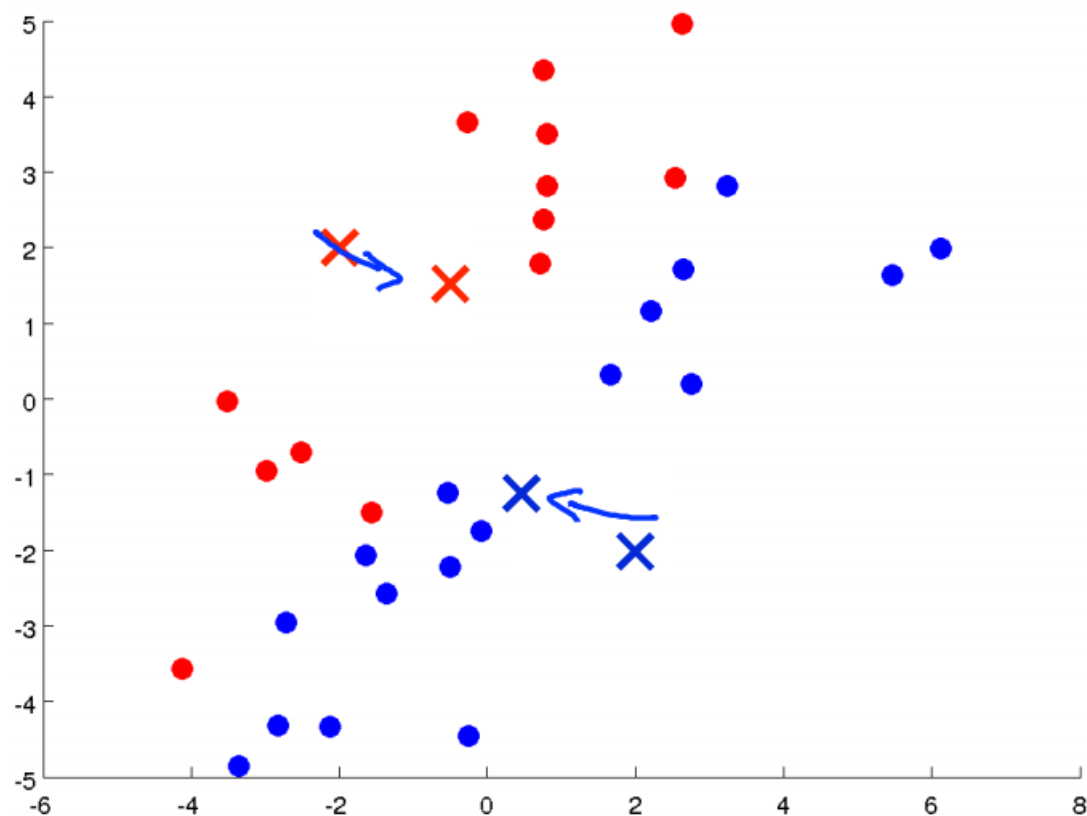
$$d(p, q) = \| p - q \|$$

$$= \sqrt{(p_1 - q_1)^2 + \dots + (p_n - q_n)^2}$$



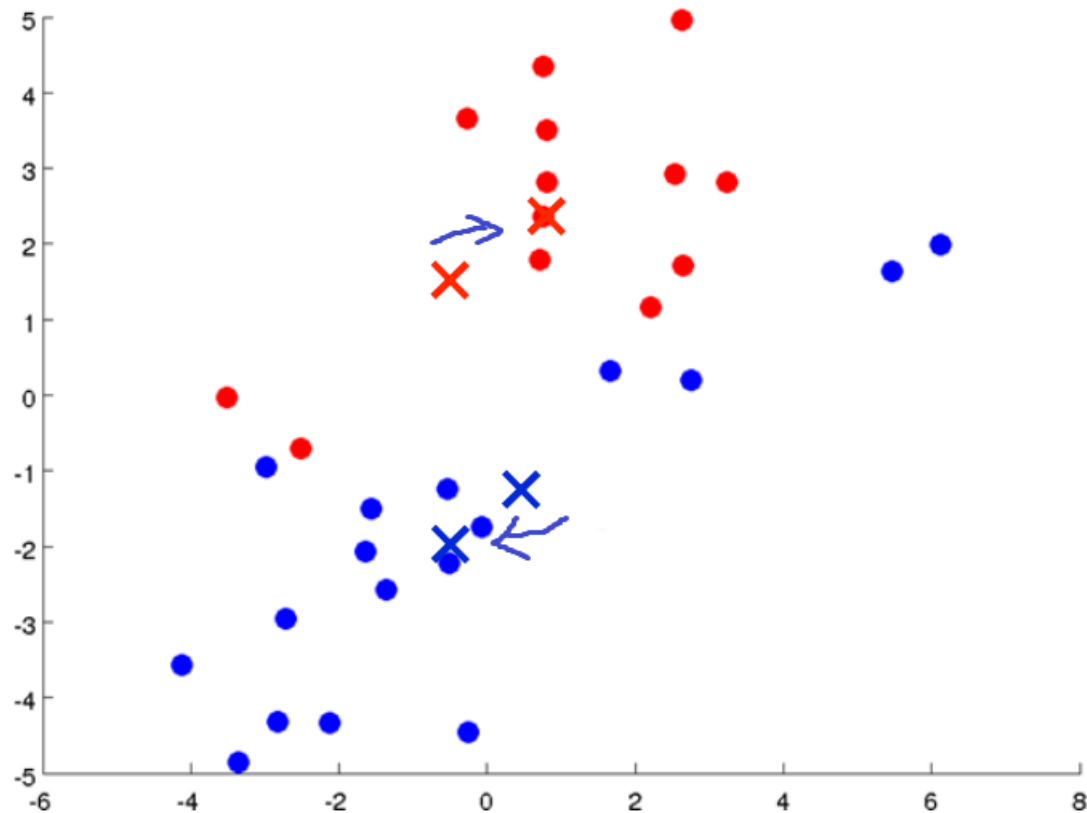
Usage of Pre-trained BERT: K-means

- 接著更新這兩個 centroid，
將其改為各個群裡資料點
的平均值
- 之後再進行一次各個資料
點與兩個 centroid 間的歐
式距離計算，重新分群
- 重複上述動作直到收斂



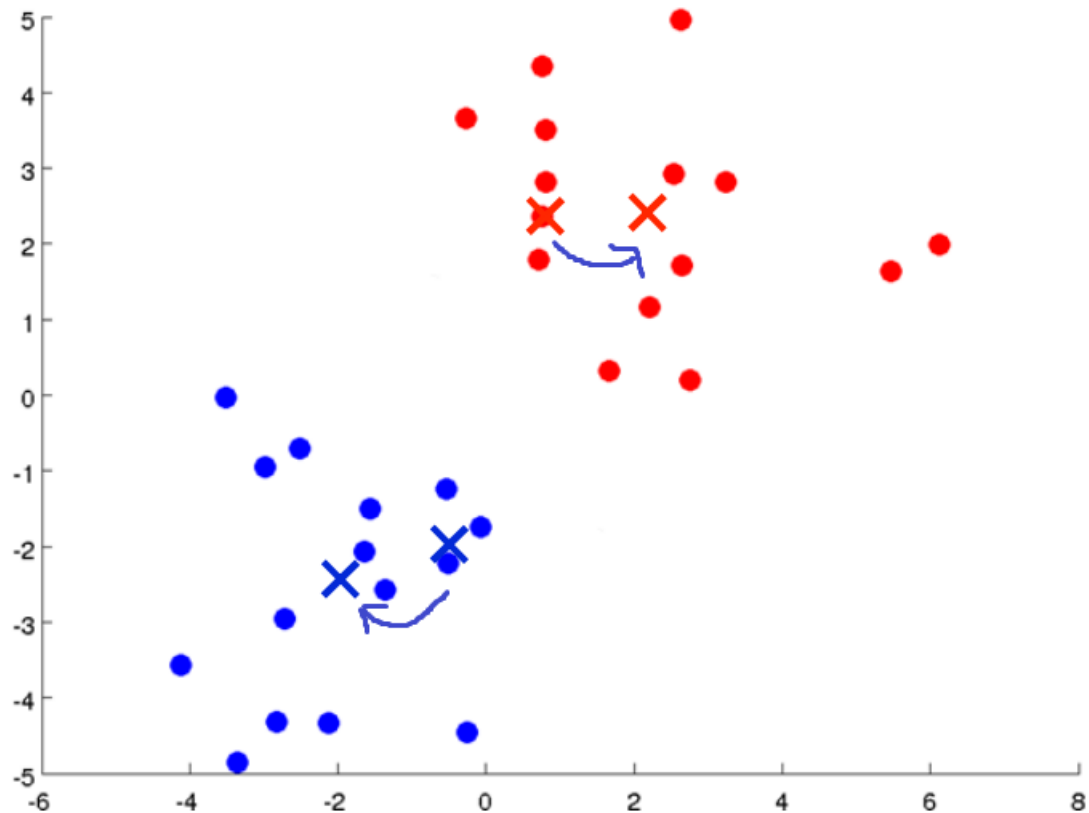
Usage of Pre-trained BERT: K-means

- 將著更新這兩個 centroid，
將其改為各個群裡資料點
的平均值
- 之後再進行一次各個資料
點與兩個 centroid 間的歐
式距離計算，重新分群
- 重複上述動作直到收斂



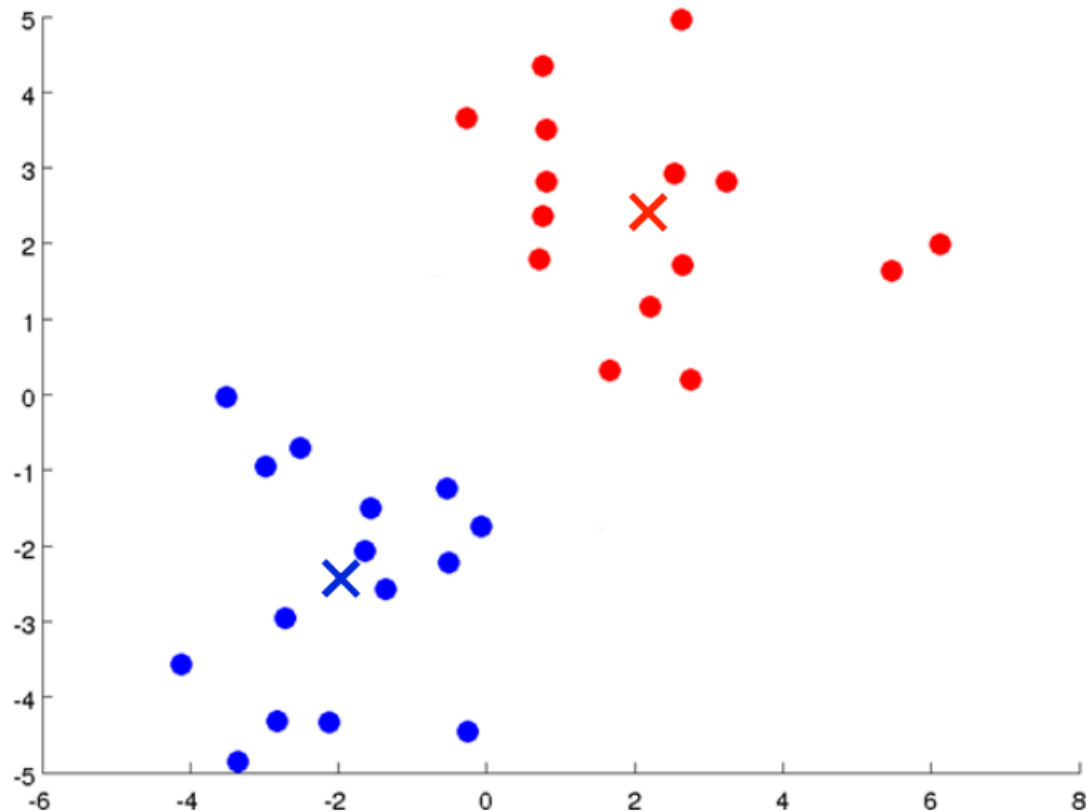
Usage of Pre-trained BERT: K-means

- 將著更新這兩個 centroid，
將其改為各個群裡資料點
的平均值
- 之後再進行一次各個資料
點與兩個 centroid 間的歐
式距離計算，重新分群
- 重複上述動作直到收斂

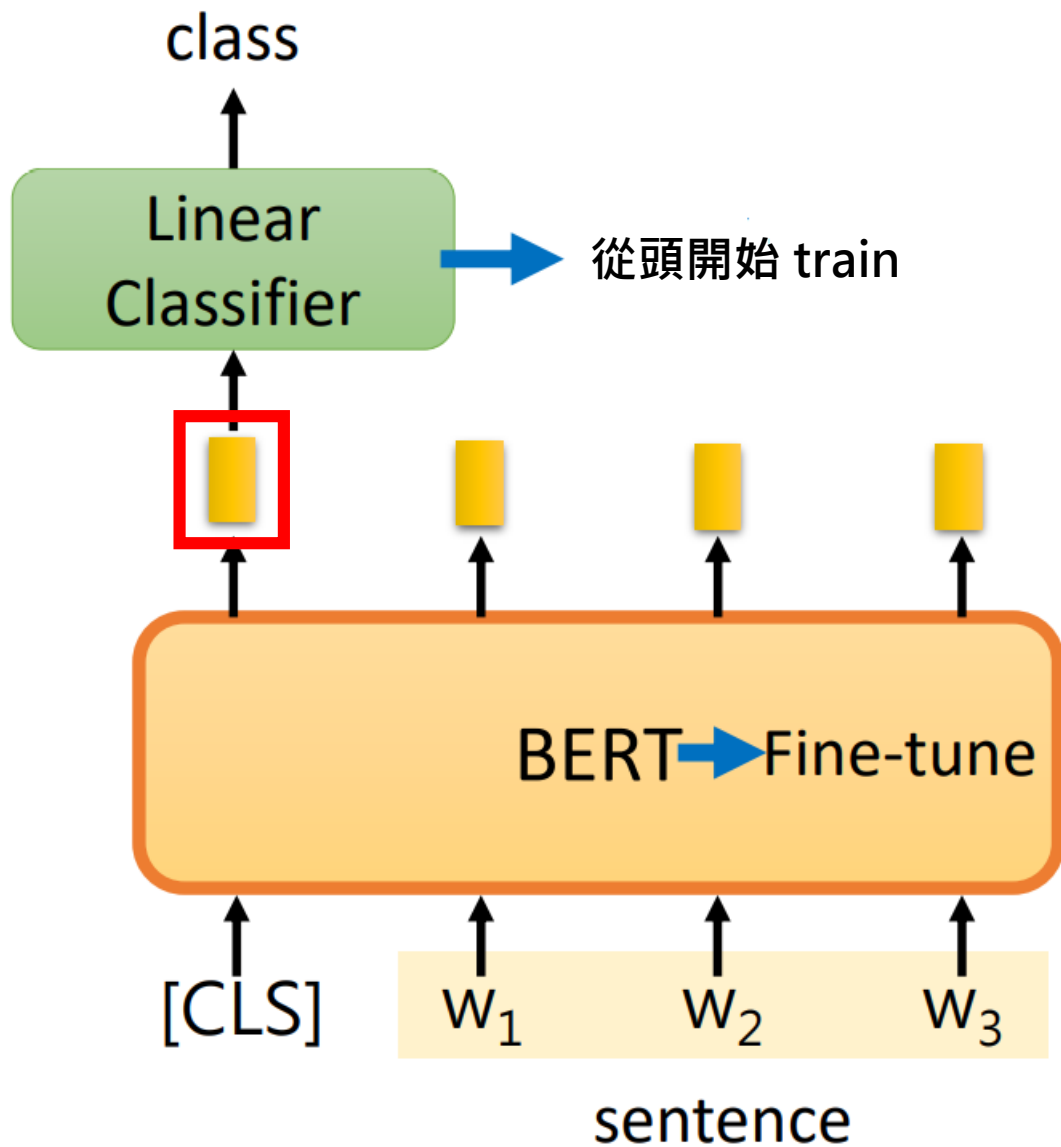


Usage of Pre-trained BERT: K-means

- 將著更新這兩個 centroid，
將其改為各個群裡資料點
的平均值
- 之後再進行一次各個資料
點與兩個 centroid 間的歐
式距離計算，重新分群
- 重複上述動作直到收斂



BERT classification



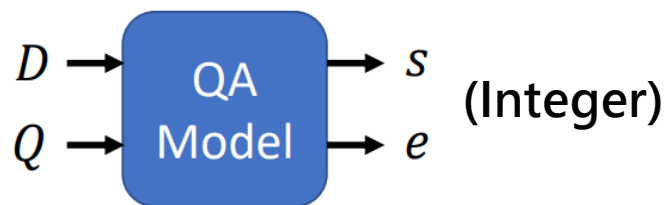
- BERT 的下游任務可以做到分類的任務。BERT 會將 input 句子的 sentence representation 丟入一個線性的分類器去做分類。
- 要注意的是，因為 BERT 的架構是預訓練好的，因此 BERT 的參數只需要微調，而新接進來的線性分類器則要從頭開始做學習。兩者在訓練時是一起訓練的。

BERT Question Answering

- BERT 的另一個下游任務是問答。我們給 BERT 一篇文章，以及一個問題，BERT 可以回答我們問題的答案，只是答案必須出現在文章裡。

Document: $D = \{d_1, d_2, \dots, d_N\}$

Query: $Q = \{q_1, q_2, \dots, q_M\}$



Answer: $A = \{d_s, \dots, d_e\}$

In meteorology, precipitation is any product of the condensation of **17** spheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **grau-pel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain **77** at **79** locations are called "showers".

What causes precipitation to fall?

gravity

$s = 17, e = 17$

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

grau-pel

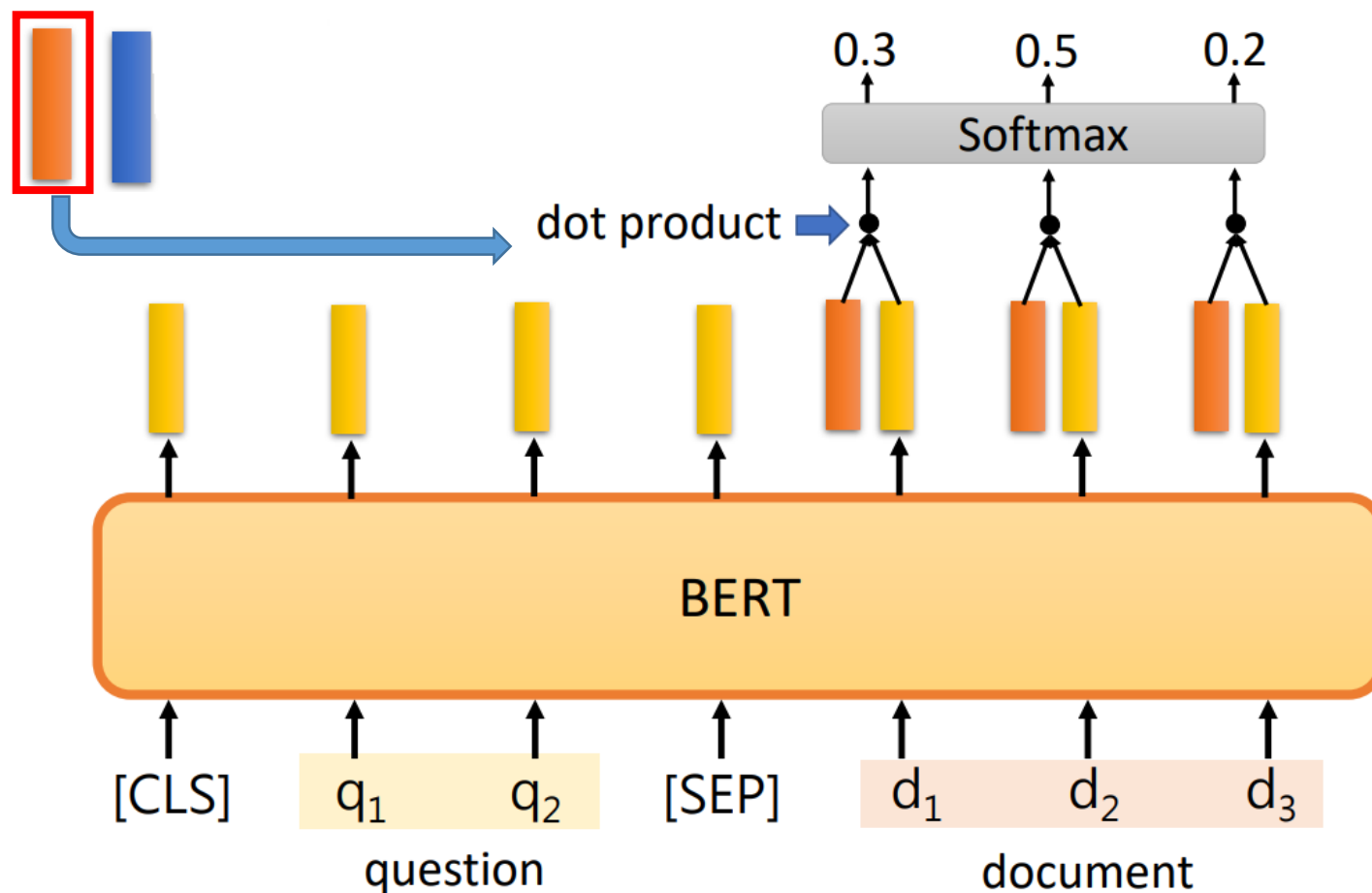
Where do water droplets collide with ice crystals to form precipitation?

within a cloud

$s = 77, e = 79$

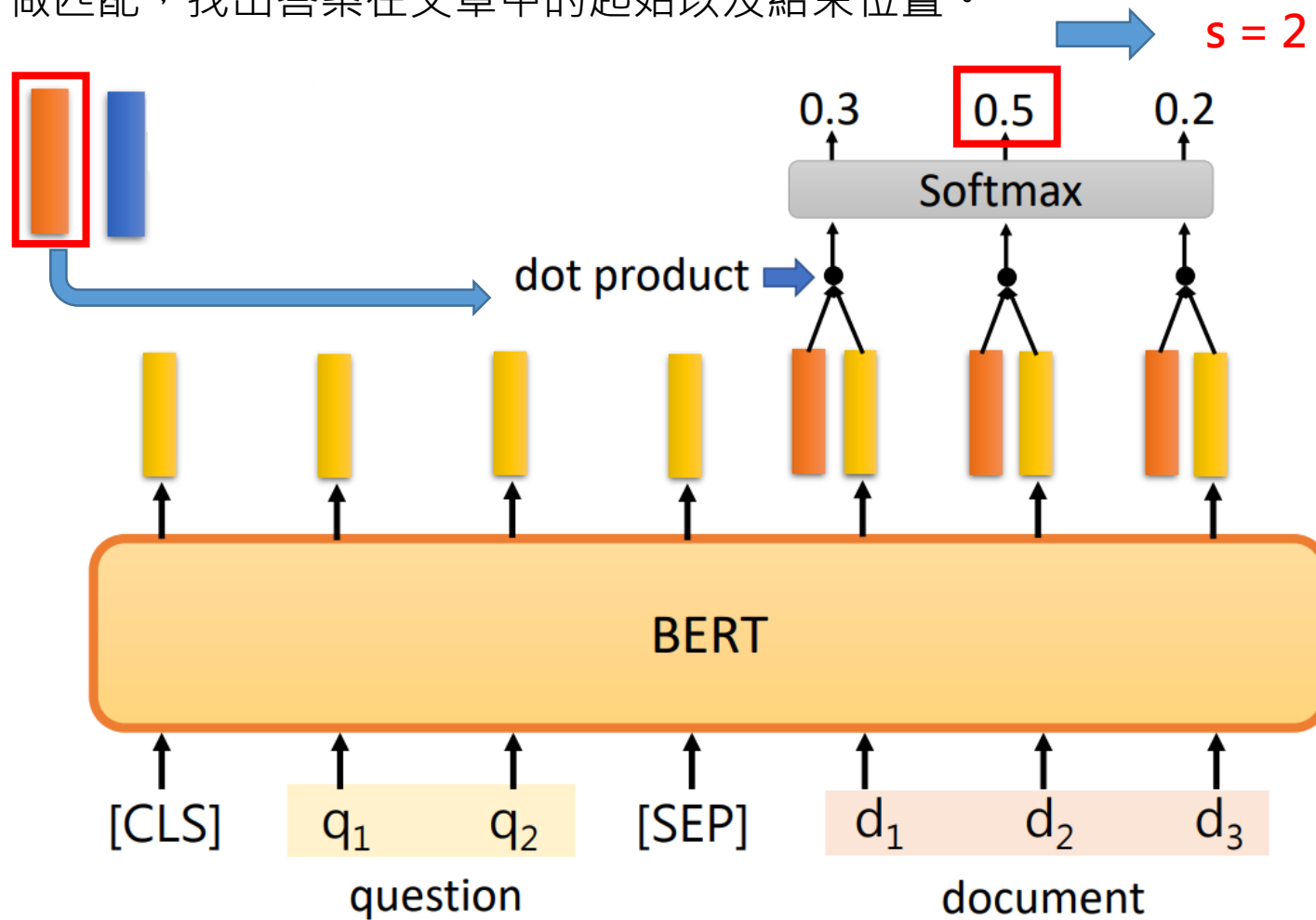
BERT Question Answering

- 重新去 learn 2個與 BERT 吐出的 word embedding 維度相同的向量。這 2個向量的作用是去與文章的每一個 word embedding 做匹配，找出答案在文章中的起始以及結束位置。



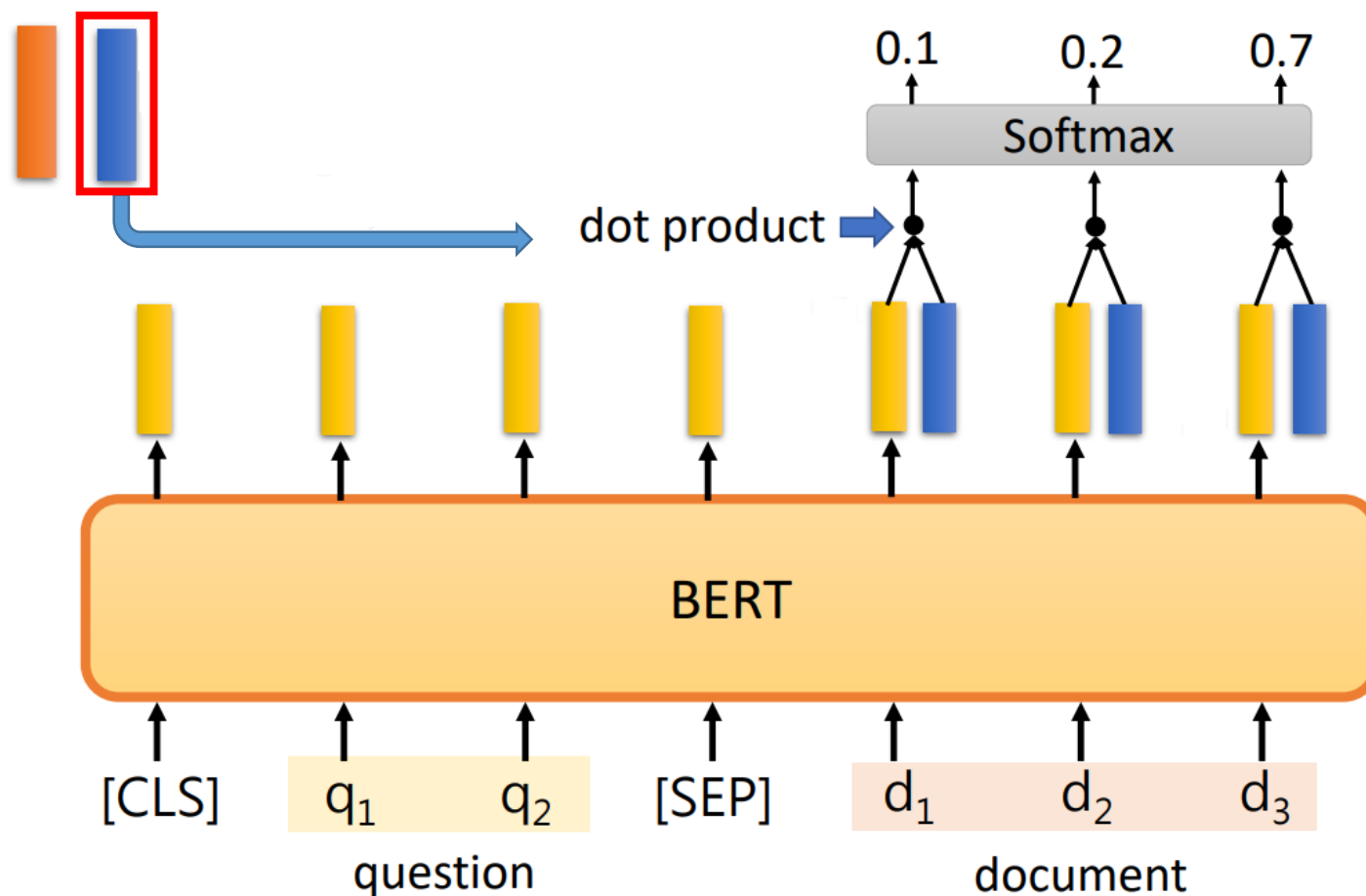
BERT Question Answering

- 重新去 learn 2個與 BERT 吐出的 word embedding 維度相同的向量。這 2個向量的作用是去與文章的每一個 word embedding 做匹配，找出答案在文章中的起始以及結束位置。



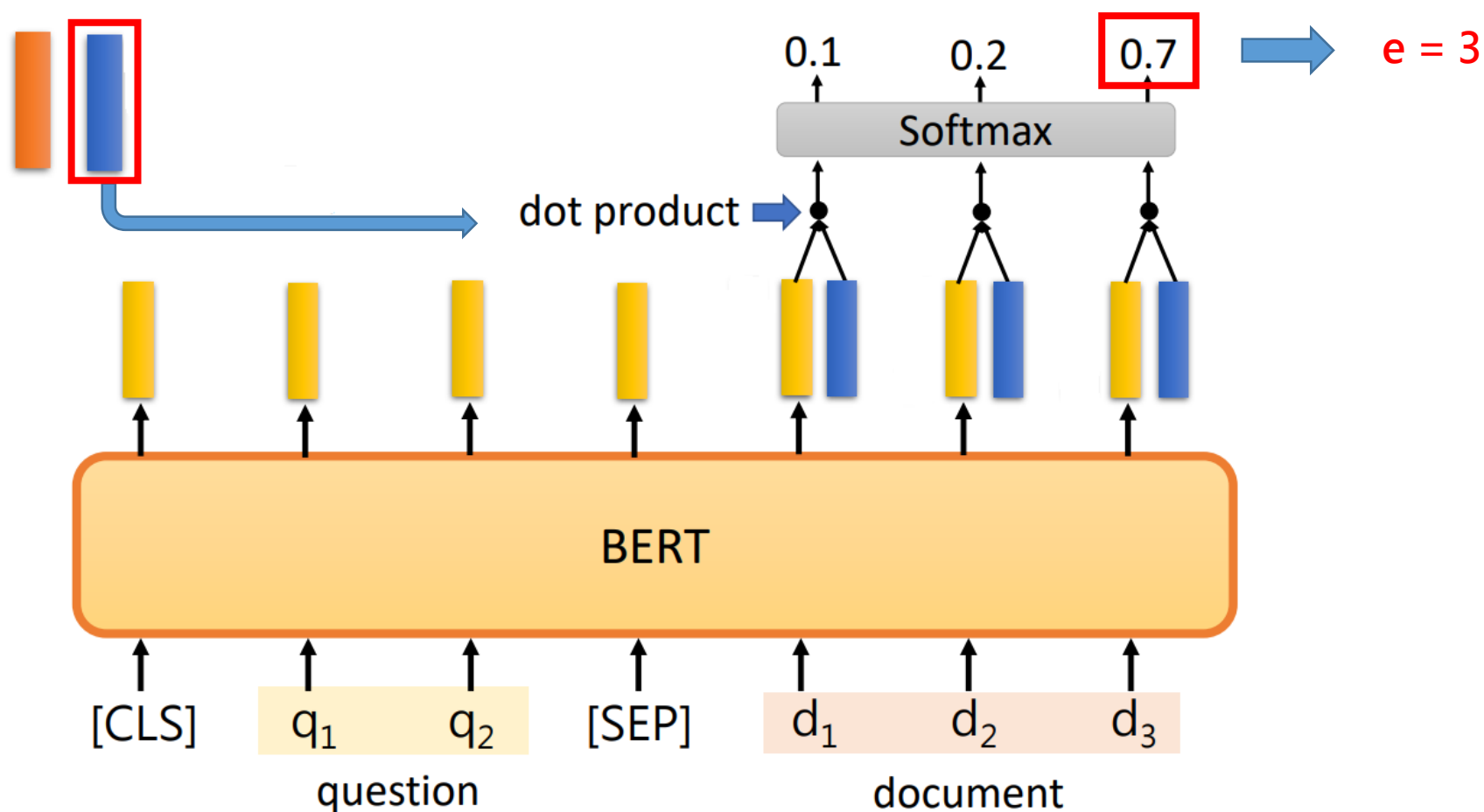
BERT Question Answering

- 重新去 learn 2個與 BERT 吐出的 word embedding 維度相同的向量。這 2個向量的作用是去與文章的每一個 word embedding 做匹配，找出答案在文章中的起始以及結束位置。



BERT Question Answering

- 重新去 learn 2個與 BERT 吐出的 word embedding 維度相同的向量。這 2個向量的作用是去與文章的每一個 word embedding 做匹配，找出答案在文章中的起始以及結束位置。



2

BERT 分類器 應用

1. 依據不同新聞主題，訓練不同模型
2. 新聞資料前處理: K-means分群，再
進行人工標籤
3. 模型訓練成果

BERT 分類任務: 依據不同主題，訓練多個模型

- 訓練資料：依據不同主題，提取其對應的資料（新聞標題）進行訓練
 - 優點：依主題訓練模型雜訊較少，較易提升模型準確率
- 資料標籤類別 (三類)：正面、中立、負面

籌碼面

外資、熱錢、
投信、法人、
大戶、籌碼、
降息、降準

需求面

需求、訂單、
大單、急單、
轉單、接單、
單量、追單、
缺貨、供應

生產面

生產、產能、
產量、產線、
產值、製造、
製程、良率、
投產

銷售面

銷售、出貨、
銷貨、銷量、
買氣、通路、
市場、後市、
後勢、展店、
市占、市佔、

營運面

營運、業績、
獲利、營利、
盈利、收益、

財報面

營收、財報、
財測、EPS、
毛利、毛利率、
淨利、純益、
盈餘、殖利率

降低手動標籤難度: K-means分群

- 由於要訓練 BERT 分類器前，必須先人工為每則新聞做標籤，新聞的數量非常多，人工大量進行標籤較為曠日廢時
- 我們將新聞標題經過預訓練的 BERT 取出資料庫裡全部新聞的 sentence embedding，並利用 K-means 進行分群。並隨機抽取各個群裡面的新聞，再進行人工標籤
- 以進行聚類，取樣後的資料去訓練表現得比未分群的結果好。其泛化程度較高，且能大量減少手動標籤的時間、數量；在相當的資料量下，較容易訓練到各種句型的標題



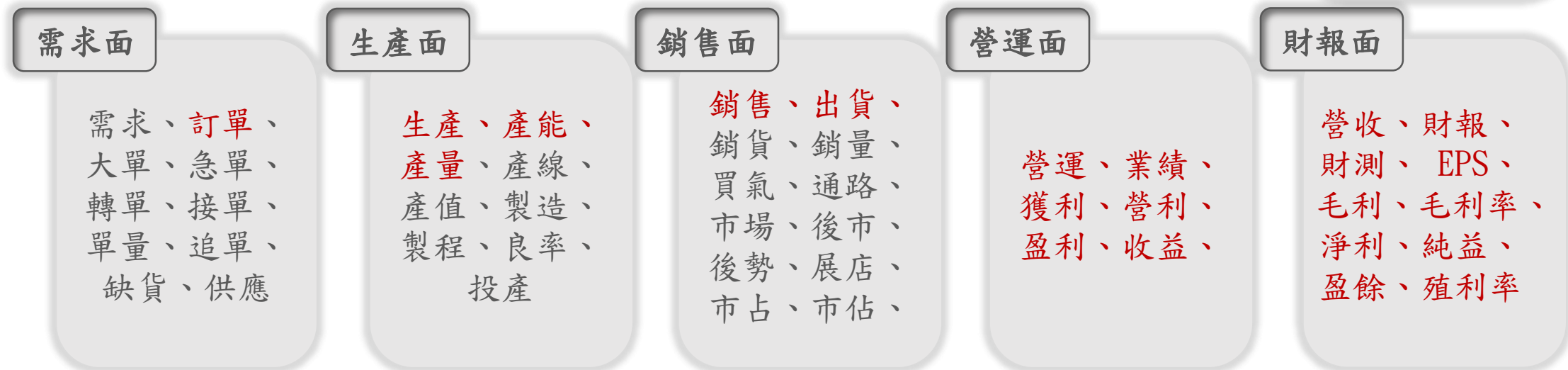
降低手動標籤難度: K-means分群

1月營收年增6%，Q1營運估落底回升	1月營收年減24%，宅經濟為Q1營運再添柴火
2月營收月減14%，復工拚Q1營運贏上季	11月營收月減9% 看好音樂串流平台發展助攻營運
1月營收年減10% Q1營運不確定性較大	1月業績小減，大陸湖北客戶佔比不到1%，今年營運穩健走揚
1月營收年減11% 第1季營運不確定性較大	1月營收年增6%！疫情衝擊「但沒有客戶銷訂單」 Q1營運估落底逐季回升

- 上圖是經過 K-means 分群後某一群的例子，可以發現確實同樣種類的新聞會被分在同一群組裡。而我們利用這個方法，隨機去取樣不同群內的新聞，再進行人工標籤。
- 經過驗證後確實可以在節省人工標籤時間下，幫助我們達到更好的訓練成果。

BERT 分類模型訓練成果

- 訓練完成模型，參數可以保留下來作使用
- 目前紅色部分為已訓練完成模型，在樣本外皆可達 90% 以上的準確率，其中財報面模型的準確率則可達 95%



BERT 分類模型訓練成果

- 訓練完成模型，參數可以保留下來作使用
- 目前紅色部分為已訓練完成模型，在樣本外皆可達 90% 以上的準確率，其中財報面模型的準確率則可達 95%

標題	類別
瑞祺電通(6416)春燕來，訂單能見度直達明年Q2，明年營運大爆發	正面
久陽訂單增 拚重返成長	正面
外資鎖定加碼台積電、面板雙虎 反手賣超鴻海逾萬張	中立
台積電7月業績減，8月起5奈米效益顯現，營收重現成長動能	中立
比亞迪口罩認證遭美否決 10億美元訂單恐告吹	負面
疫情衝擊！日本造船訂單腰斬、積壓訂單1997年來新低	負面

3

BERT QA 應用

1. BERT QA: 擷取新聞用語的關鍵情緒
詞彙
2. BERT QA: 新聞資料前處理
3. BERT QA: 模型訓練成果

BERT 分類任務的缺點

- 無法體現一則新聞標題正面、負面程度上的差異: 交易員提出，在短期內要以新聞標題為一交易策略指標的話，不僅僅需要知道一則新聞是正面、負面或是中立，更需要得到每則新聞有多正面或有多負面。

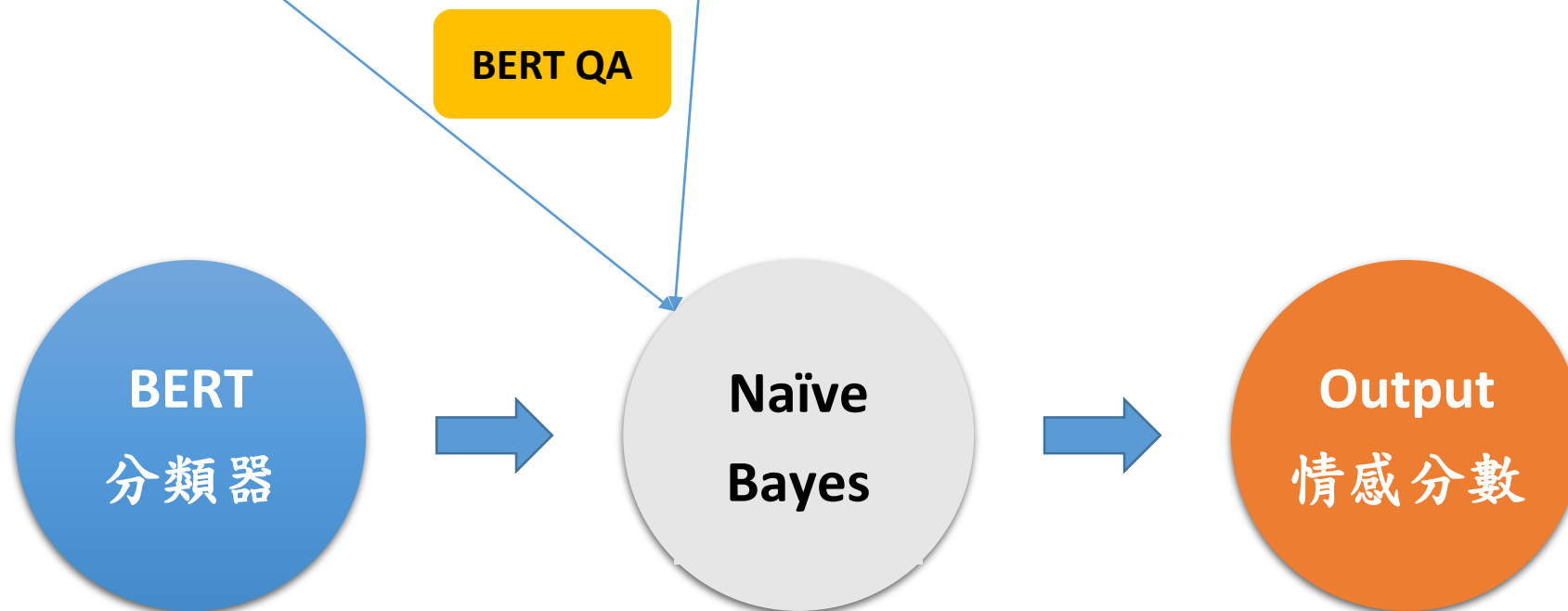
標題	類別
瑞祺電通(6416)春燕來，訂單能見度直達明年Q2，明年營運大爆發	正面
久陽訂單增 拚重返成長	正面



BERT 分類任務的缺點

- 我想到的方法是透過擷取每則新聞的**關鍵情緒詞彙**，並透過 Naïve Bayes 模型去對這則新聞做評分

標題	類別
瑞祺電通(6416) 春燕來 ，訂單能見度直達明年Q2，明年營運 大爆發	正面
久陽訂單 增 拚重返成長	正面



BERT QA: 新聞資料前處理

- 人工對新聞資料做標籤

文章

問題

答案

▲個股：上半年獲利看俏+現金減資45.44795%及配息1.5元，國揚盤中股價漲停	獲利表現如何?#股價表現如何?	看俏#漲停
△個股：LuLu上修第四季獲利預估，股價創新天價，儒鴻(1476)受惠，今年展望佳	獲利表現如何?#股價表現如何?	上修#創新天價
△個股：台驊上半年獲利可期，股價帶量突破大聯大收購價，盤中站上30元關卡	獲利表現如何?#股價表現如何?	可期#帶量突破
△個股：晶技(3042)Q1獲利佳，法人估Q2營收季增10-15%，早盤股價強攻漲停	獲利表現如何?#營收表現如何?#股價表現如何?	佳#季增#強攻漲停
△個股：板卡Q2淡季不淡，微星、技嘉、華擎4月業績攻高，股價勁揚	業績表現如何?#股價表現如何?	攻高#勁揚

- 這次我標籤了 800 多筆訓練資料，共涵蓋「營運、訂單、營收、需求、業績、EPS、出貨、股價、獲利、外資」等 10 種主題。
- 在做標籤前，一樣有先經過 K-means 分群並抽樣。

BERT QA: 訓練流程

文章

答案位置 → start: 9、end: 10

問題

答案

▲個股：上半年獲利看俏，現金減資45.44795%及配息1.5元，國揚盤中股價漲停
△個股：LuLu上修第四季獲利預估，股價創新天價，儒鴻(1476)受惠，今年展望佳
△個股：台驊上半年獲利可期，股價帶量突破大聯大收購價，盤中站上30元關卡
△個股：晶技(3042)Q1獲利佳，法人估Q2營收季增10-15%，早盤股價強攻漲停
△個股：板卡Q2淡季不淡，微星、技嘉、華擎4月業績攻高，股價勁揚

獲利表現如何?#股價表現如何?
獲利表現如何?#股價表現如何?
獲利表現如何?#股價表現如何?
獲利表現如何?#營收表現如何?#股價表現如何?
業績表現如何?#股價表現如何?

看俏#漲停
上修#創新天價
可期#帶量突破
佳#季增#強攻漲停
攻高#勁揚

讀文章

將標題轉為 BERT
看得懂的形式
(Embedding)

接收問題

將問題轉為 BERT
看得懂的形式
(Embedding)

回答

BERT 學習出答案在文章中的
起始、結束位置

BERT QA: 模型訓練成果

- 模型訓練 - 運算處理器：Colab TPU
- Server上預測速度：約 30個問題 / 秒
- 主題：

營運、訂單、營收、需求、業績、EPS、出貨、股價、獲利、外資

測試集準確率：**0.90**

標題	answer
瑞祺電通(6416)春燕來，訂單能見度直達明年Q2，明年營運大爆發	春燕來#大爆發
久陽訂單增 拚重返成長	增
比亞迪口罩認證遭美否決 10億美元訂單恐告吹	恐告吹
疫情衝擊！日本造船訂單腰斬、積壓訂單1997年來新低	腰斬

4

情感分數 Naïve Bayes

1. NB: 模型介紹
2. NB: 情感分析器建構流程
3. NB: 情感分析器訓練成果

NB: 模型介紹

➤ Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

➤ Bayes Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{P(A \cap B)}{P(A)} * P(A)}{P(B)} = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|A^c) * P(A^c)}$$

➤ Naïve Bayes (for binary classification)

$$P(y = 0|X_1, X_2, \dots, X_n) = \frac{P(X_1, X_2, \dots, X_n|y = 0) * P(y = 0)}{P(X_1, X_2, \dots, X_n|y = 0) * P(y = 0) + P(X_1, X_2, \dots, X_n|y = 1) * P(y = 1)}$$

Naive Bayes suppose X_1, X_2, \dots, X_n are independent, then we have

$$P(X_1, X_2, \dots, X_n|y = 0) = P(X_1|y = 0) * P(X_2|y = 0) * \dots * P(X_n|y = 0)$$

NB: 情感分析應用

➤ $y = 1$ (正面詞彙)

上漲

x3

看好

x2

滿載

x5

➤ $y = 0$ (負面詞彙)

下跌

x5

惡化

x3

看空

x2

NB: 情感分析應用

➤ $y = 1$ (正面詞彙)

上漲
x3

看好
x2

滿載
x5

$$P(y = 1 | \text{上漲}) = \frac{P(\text{上漲} | y = 1) * p(y = 1)}{P(\text{上漲} | y = 1) * p(y = 1) + P(\text{上漲} | y = 0) * p(y = 0)} = 1$$

➤ $y = 0$ (負面詞彙)

下跌
x5

惡化
x3

看空
x2

$$P(y = 1 | \text{下跌}) = \frac{P(\text{下跌} | y = 1) * p(y = 1)}{P(\text{下跌} | y = 1) * p(y = 1) + P(\text{下跌} | y = 0) * p(y = 0)} = 0$$

NB: 情感分析應用

➤ $y = 1$ (正面詞彙)

上漲 x3	看好 x2	滿載 x5
----------	----------	----------

$$P(y = 1 | \text{上漲}) = \frac{P(\text{上漲} | y = 1) * p(y = 1)}{P(\text{上漲} | y = 1) + P(\text{上漲} | y = 0) * p(y = 0)} = 1$$

➤ $y = 0$ (負面詞彙)

下跌 x5	惡化 x3	看空 x2
----------	----------	----------

$$P(y = 1 | \text{下跌}) = \frac{P(\text{下跌} | y = 1) * p(y = 1)}{P(\text{下跌} | y = 1) + P(\text{下跌} | y = 0) * p(y = 0)} = 0$$

NB: 情感分析應用

➤ $y = 1$ (正面詞彙)

上漲 x4	看好 x3	滿載 x6	下跌 x1	惡化 x1	看空 x1
----------	----------	----------	----------	----------	----------

➤ $y = 0$ (負面詞彙)

下跌 x6	惡化 x4	看空 x3	上漲 x1	看好 x1	滿載 x1
----------	----------	----------	----------	----------	----------

$$P(y = 1 | \text{上漲})$$

$$= \frac{P(\text{上漲} | y = 1) * p(y = 1)}{P(\text{上漲} | y = 1) * p(y = 1) + P(\text{上漲} | y = 0) * p(y = 0)}$$

$$= \frac{\frac{4}{16} * \frac{16}{32}}{\frac{4}{16} * \frac{16}{32} + \frac{1}{16} * \frac{16}{32}} = 0.8$$

$$P(y = 1 | \text{下跌})$$

$$= \frac{P(\text{下跌} | y = 1) * p(y = 1)}{P(\text{下跌} | y = 1) * p(y = 1) + P(\text{下跌} | y = 0) * p(y = 0)}$$

$$= \frac{\frac{1}{16} * \frac{16}{32}}{\frac{1}{16} * \frac{16}{32} + \frac{6}{16} * \frac{16}{32}} = 0.143$$

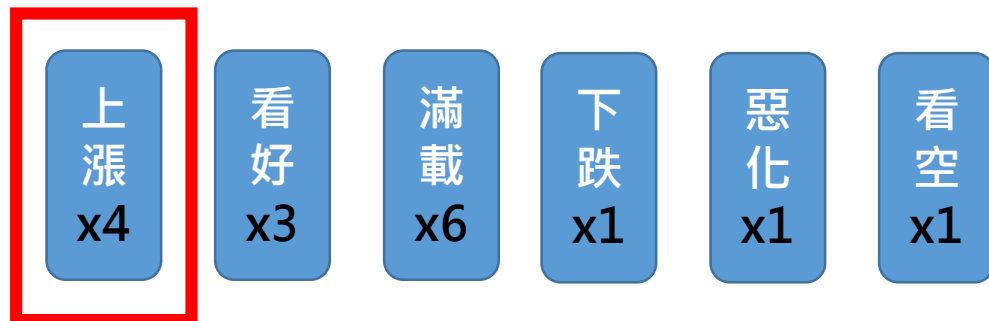
NB: 情感分析應用

$$\begin{aligned}P(y = 1 | \text{上漲, 看好}) &= \frac{P(\text{上漲, 看好} | y = 1) * p(y = 1)}{P(\text{上漲, 看好} | y = 1) * p(y = 1) + P(\text{上漲, 看好} | y = 0) * p(y = 0)} \\&= \frac{P(\text{上漲} | y = 1) * P(\text{看好} | y = 1) * p(y = 1)}{P(\text{上漲} | y = 1) * P(\text{看好} | y = 1) * p(y = 1) + P(\text{上漲} | y = 0) * P(\text{看好} | y = 0) * p(y = 0)} \\&= \frac{\frac{4}{16} * \frac{3}{16} * \frac{16}{32}}{\frac{4}{16} * \frac{3}{16} * \frac{16}{32} + \frac{1}{16} * \frac{1}{16} * \frac{16}{32}} = \mathbf{0.923}\end{aligned}$$

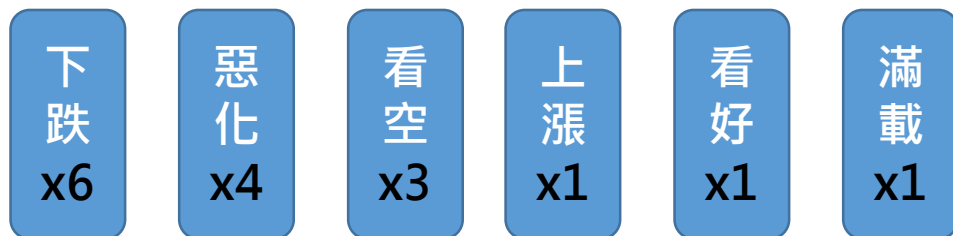
$$P(y = 1 | \text{上漲, 看空}) = \frac{\frac{4}{16} * \frac{1}{16} * \frac{16}{32}}{\frac{4}{16} * \frac{1}{16} * \frac{16}{32} + \frac{1}{16} * \frac{3}{16} * \frac{16}{32}} = \mathbf{0.57}$$

NB: 情感分析應用

➤ $y = 1$ (正面詞彙)

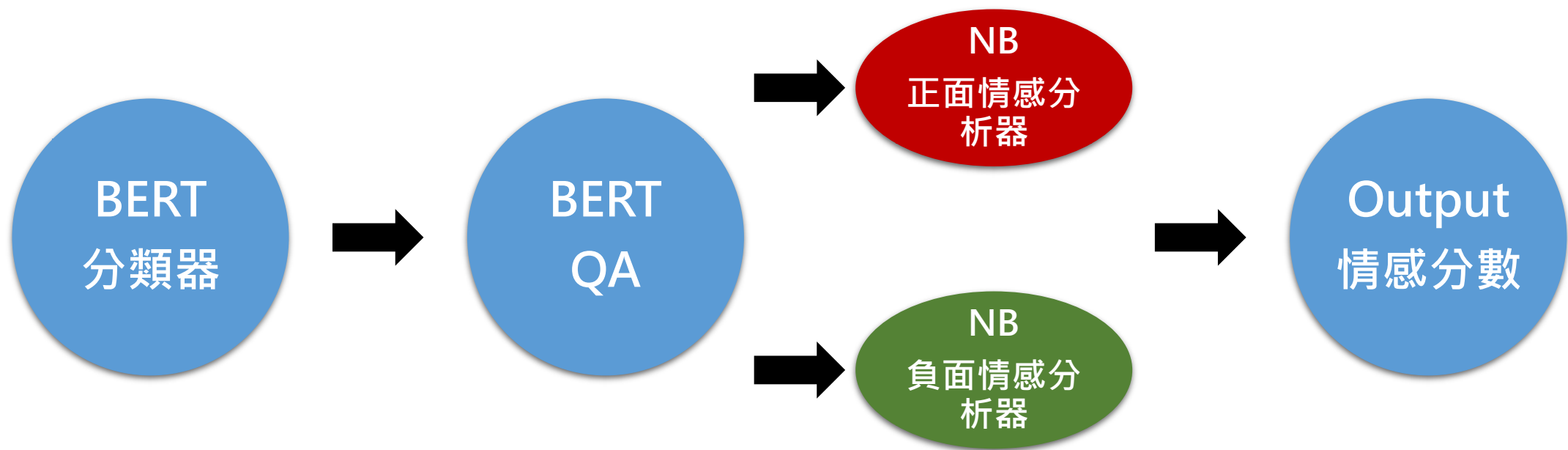


➤ $y = 0$ (負面詞彙)

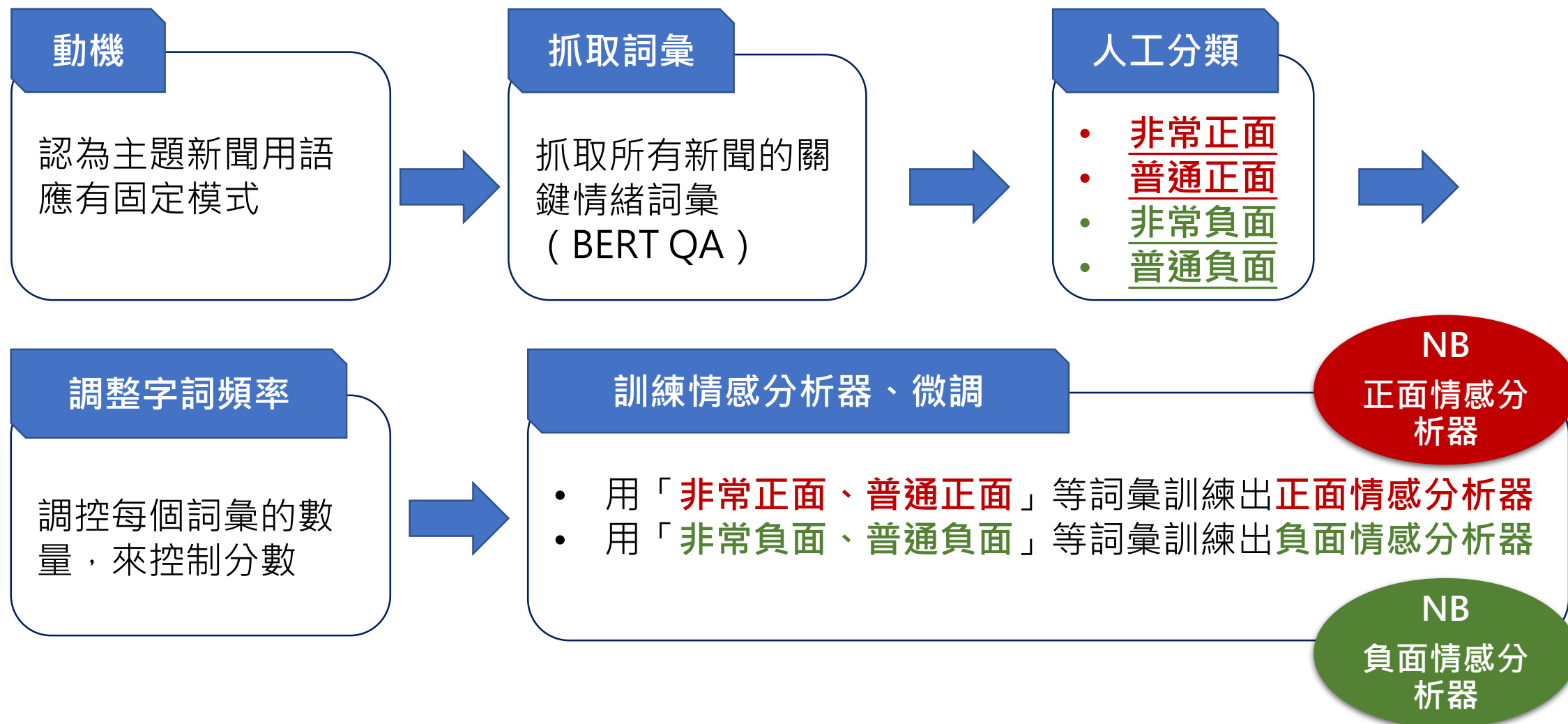


- 假如我們希望「上漲」這個詞彙的情感分數更高，我們可以在正面詞彙字典裡，將其數量增加
- 透過 Naïve Bayes 情感分析器，我們能夠透過調整詞頻，來對每個字詞的分數做一定的掌控
- 再者，我們假設新聞所運用的情緒詞彙用語有限，因此我們可以用 BERT QA 去抓取出所有的新聞情緒詞彙，並訓練自己的情感分析器

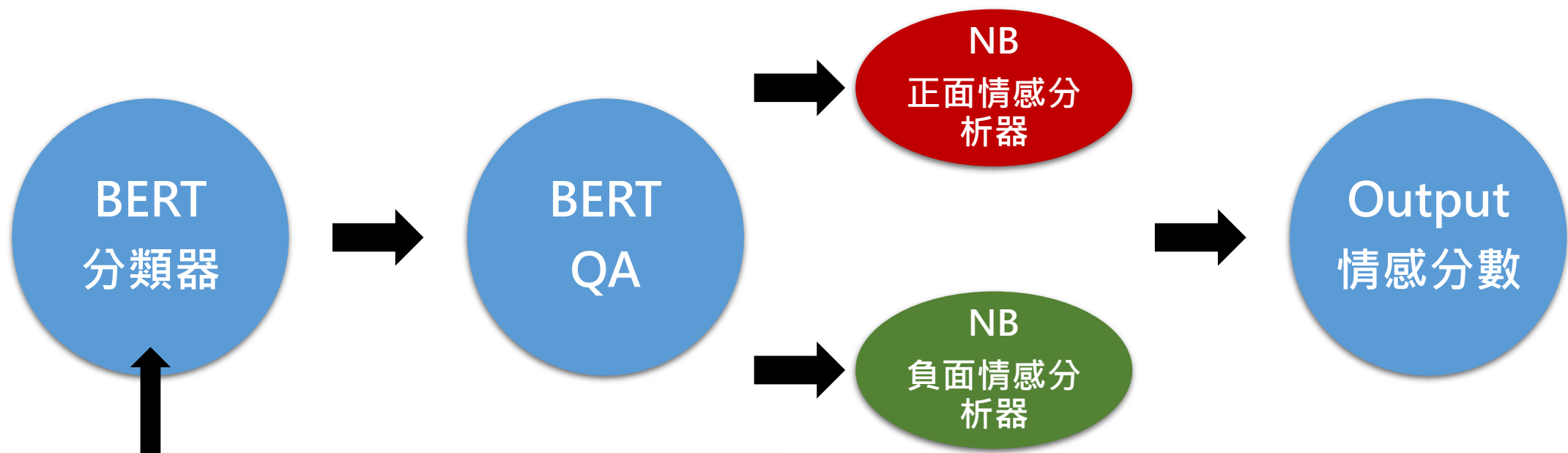
NB: 情感分析器建構流程



NB: 情感分析器建構流程

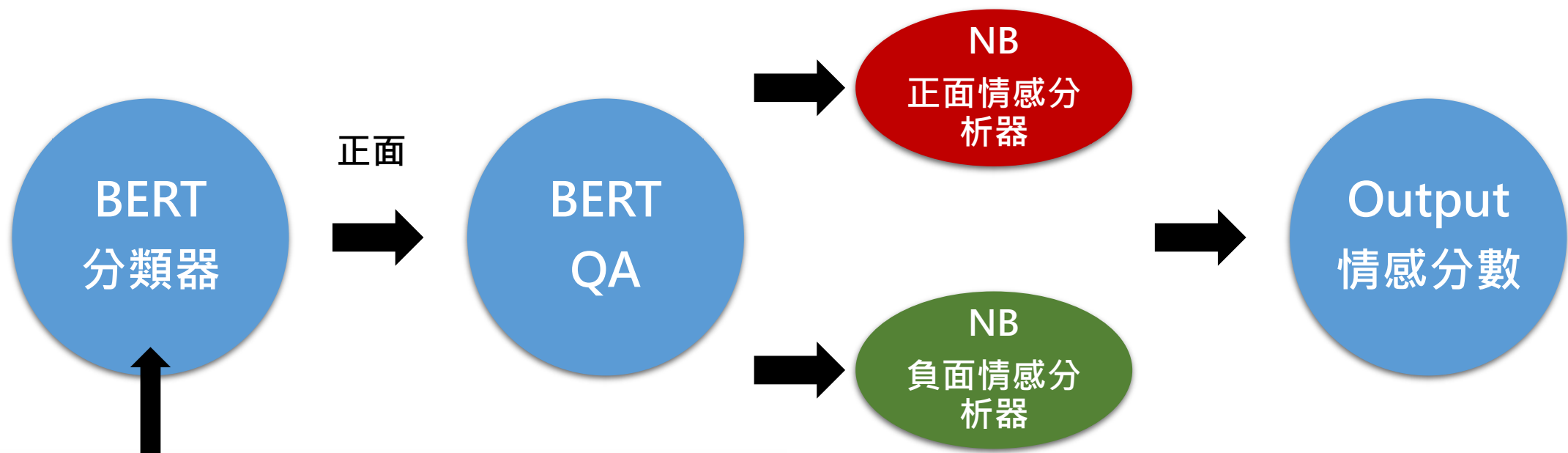


NB: 情感分析器建構流程



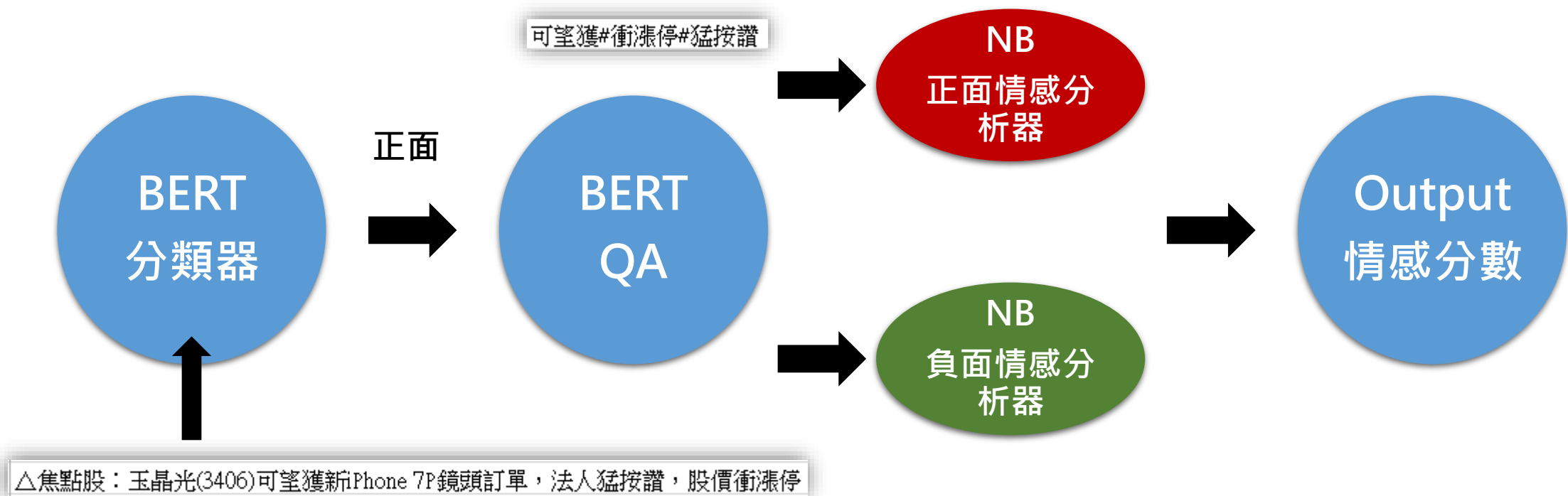
△焦點股：玉晶光(3406)可望獲新iPhone 7P鏡頭訂單，法人猛按讚，股價衝漲停

NB: 情感分析器建構流程

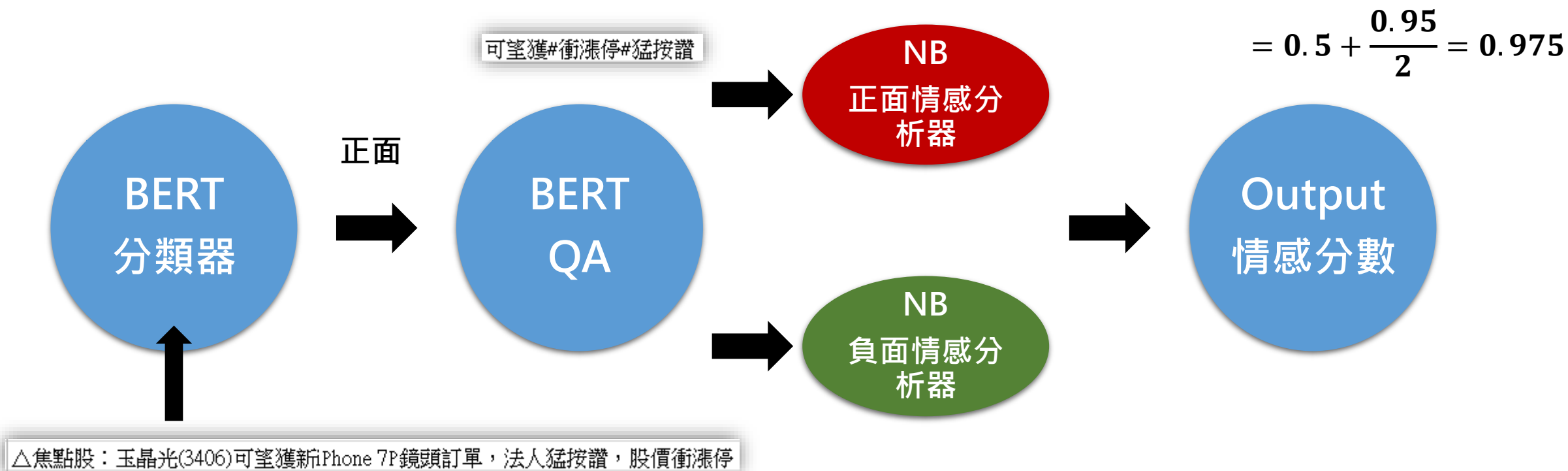


△焦點股：玉晶光(3406)可望獲新iPhone 7P鏡頭訂單，法人猛按讚，股價衝漲停

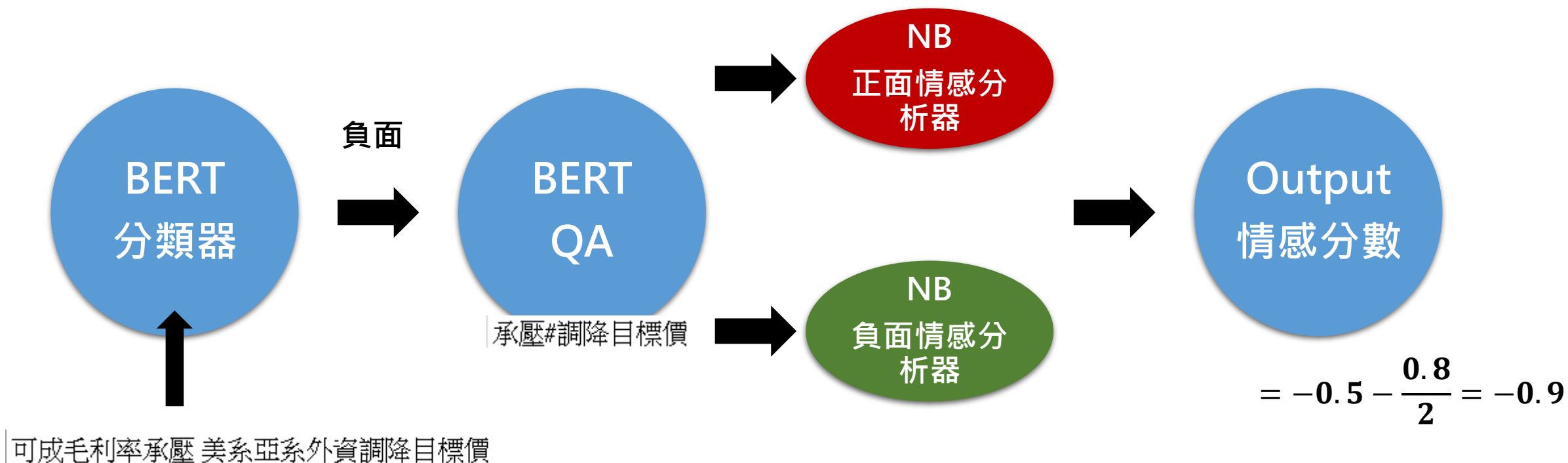
NB: 情感分析器建構流程



NB: 情感分析器建構流程



NB: 情感分析器建構流程



NB: 情感分析器訓練成果

情感分數計算方法

- BERT 分類器判斷為**正面**: $0.5 + \frac{NB\ score}{2}$
- BERT 分類器判斷為**負面**: $-0.5 - \frac{NB\ score}{2}$

標題	類別	情感分數	總分
瑞祺電通(6416)春燕來，訂單能見度直達明年Q2，明年營運大爆發	正面	0.99	0.99
和大(1536)明年特斯拉訂單增逾60%，估全年營收75億元起跳創新高	正面	0.73	0.86
久陽訂單增拚重返成長	正面	0.09	0.54
Google高階伺服器板爆材料瑕疵 傳訂單已轉向日本松電工 台耀痛失大單	負面	0.98	- 0.99
訂單能見度不佳，外資下修大江目標價	負面	0.48	- 0.74
〈大亞展望〉疫情干擾本業Q2訂單能見度低 營運保守看	負面	0.008	- 0.50