

1. 以下何種方法所產生的字詞向量的維度是可解釋的?

- A. CBOW
- B. Skip gram
- C. SVD
- D. TF-IDF

ANS: D, TF-IDF 每一維度是字詞

2. 以下何種方法不是使用神經網路來產生字詞向量?

- A. SVD
- B. Skip gram
- C. CBOW
- D. Glove

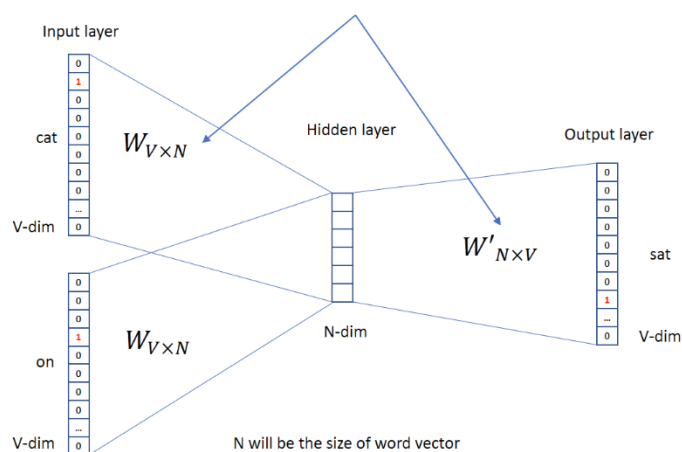
ANS : A , SVD 採用矩陣分解法來產生字詞向量

3. 相對於傳統產生文字向量的方法(如 SVD) , 那個不是使用神經網路來產生字詞向量的好處?

- A. 有考慮文字的鄰近性
- B. 維度較低
- C. 新的訓練文件可以隨時加入, 不用重新訓練
- D. 也可以同時訓練文件的向量

ANS : B , 傳統方法和神經網路在產生字詞向量的維度數沒有差別。

考慮以下的 CBOW 神經網路圖



以下為 $W_{V \times N}^T$

V

0.2	1.2	0.8	0.1	2.1	3.1	2.1	1.4	...
3.1	0.2	0.3	1.2	0.2	1.2	4.2	0.2	...
2.1	3.6	1.2	3.1	0.1	0.2	1.3	4.3	...
1.5	2.3	9.2	0.1	0.5	3.5	1.4	1.3	...
2.3	1.8	9.1	0.4	0.2	1.5	8.4	0.3	...
5.4	1.2	0.3	4.3	1.2	3.2	5.4	4.3	...
3.2	1.5	4.3	1.0	2.3	8.4	1.2	2.3	...
...
3.4	2.1	0.2	2.3	1.3	3.2	1.2	2.7	...

N

4. “cat”的第一個維度值為

- A. 0.2
- B. 1.2
- C. 0.8
- D. 0.1

ANS：B，cat 在第 2 行（column）

5. “on”的第一個維度值為

- A. 0.2
- B. 1.2
- C. 0.8
- D. 0.1

ANS：D，on 在第 4 行（column）

6. “sat”的第一個維度值為

- A. 0.1
- B. 2.1
- C. 3.1
- D. 1.4

ANS：D，sat 在第 8 行（column）

7. 以下何種不是 word2vec 的應用？

- A. 判斷詞跟詞的相似程度
- B. 查詢詞自動擴充
- C. 判斷詞跟詞間的關係

D. 找出同音字詞

ANS：D，word2vec 沒有處理字詞發音

8. 以下何者不是 Transformer Model 的特色？

- A. 參數量很大
- B. 分成 Encoder 和 Decoder
- C. BERT 是 Encoder
- D. ChatGPT 包含 Encoder 和 decoder

ANS：D，ChatGPT 是 decoder-only 架構

9. 跟 seq2seq Model 比起來，Transformer Model 有何優點？

- A. 較能注意到長文的前後文關係
- B. 較能使用 GPU 的平行處理
- C. 可以使用大量文本訓練出一個模型供後需下游任務來使用
- D. 以上皆是

ANS：D，都是。

10. 一下那個 embeddings 不是 Transformer Model 訓練出來的

- A. Skip-gram
- B. FastText
- C. Bert
- D. Cohere

ANS：A，Skip-gram 使用簡單的 2 層式神經網路架構設計的。