Representation of Image, Text and Signals

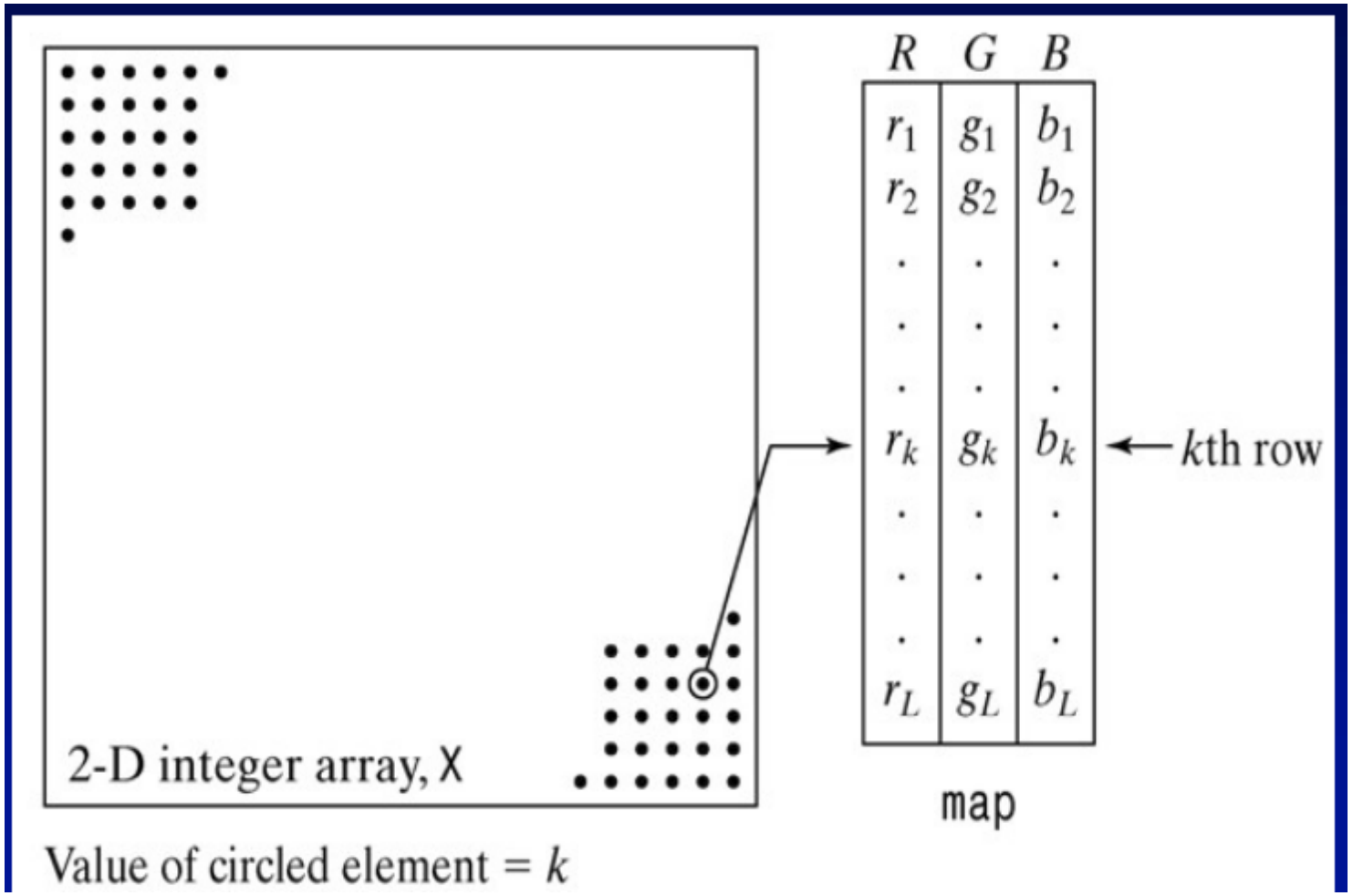
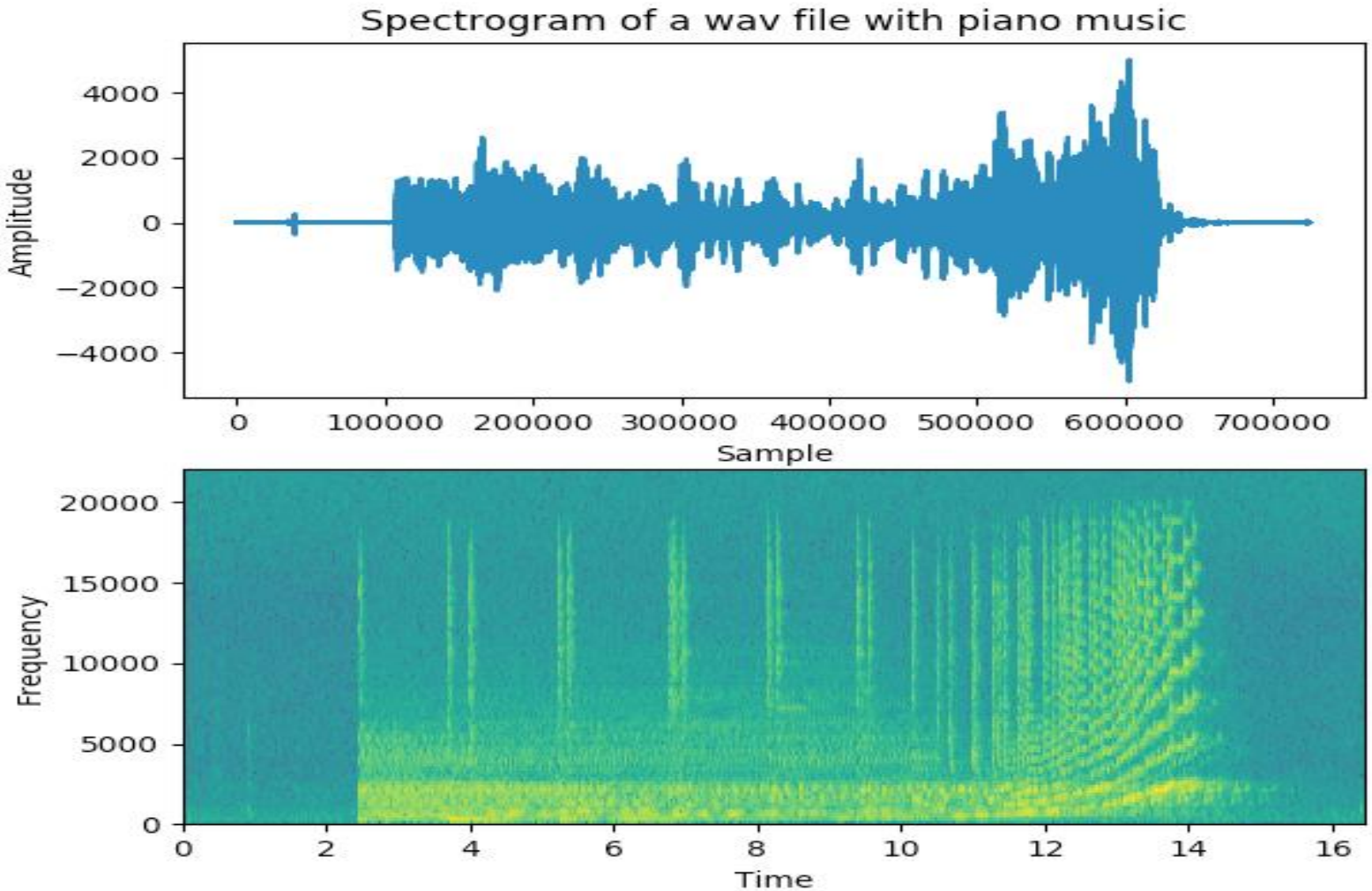


Image to Matrix representation



Signal to Matrix representation

Text to Matrix representation

Sentece 1: We are in DCS.
Sentece 2: DCS is in Vidyavardhaka.
Sentece 3: Vidyavardhaka is in Mysore.

Vocabulary/Dictionary = Vidyavardhaka, are, We, in, DCS, is, Mysore

	Vidyavardhaka	are	We	in	DCS	is	Mysore
Sentence 1	0	1	1	1	1	0	0
Sentence 2	1	0	0	1	1	1	0
Sentence 3	1	0	0	1	0	1	1

1. Bag-of-words
2. Term document matrix (TDM)
3. Term-frequency inverse document frequency (TFIDF)
4. TDM+SVD
5. TDM+NMF
6. TFIDF+SVD
7. TFIDF+NMF
8. Doc2vec
9. Word embedding (Skip gram, CBOW, and glovec)
10. fastText
11. BERT (Bidirectional Encoder Representations from Transformers)

Word embedding (Skip gram)

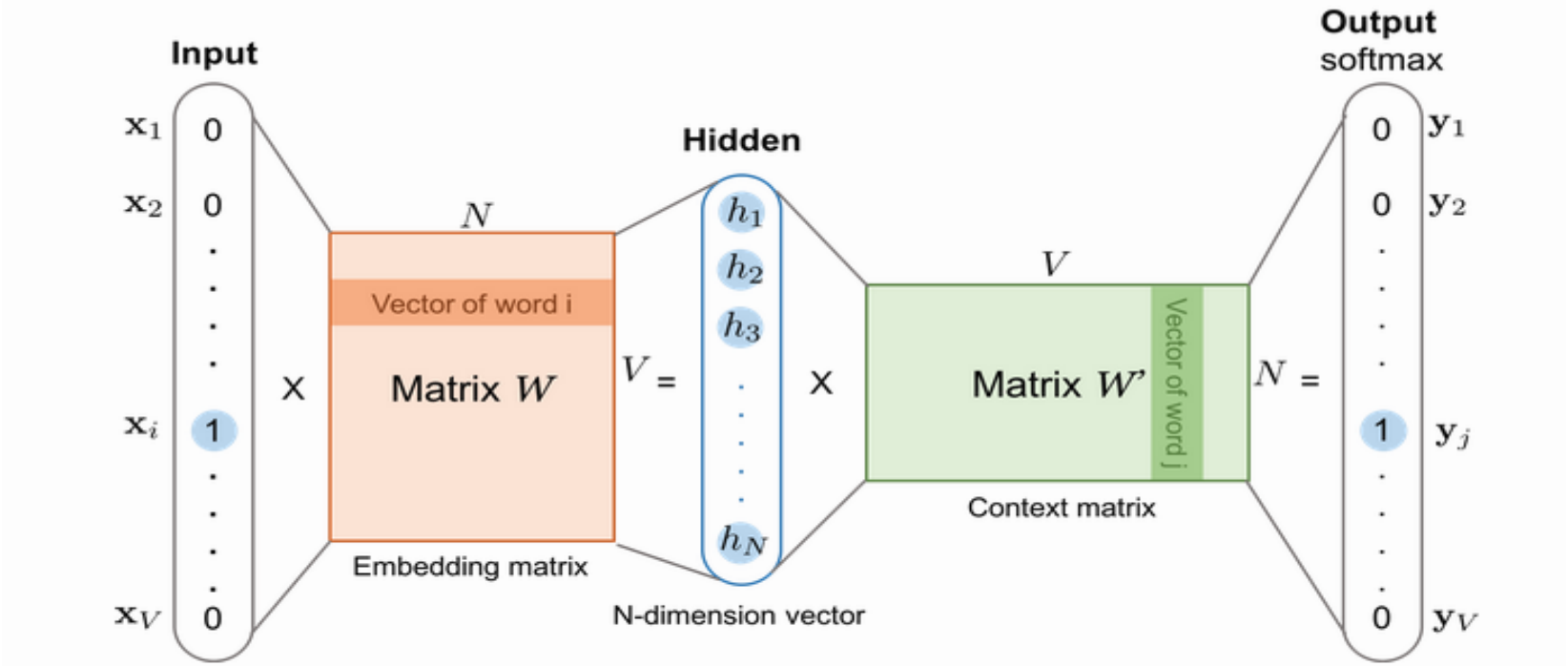


Fig. 1. The skip-gram model. Both the input vector $\mathbf{x}$ *and the output* $\mathbf{y}$ *are one-hot encoded word representations. The hidden layer is the word embedding of size* N.

Word embedding (CBOW)

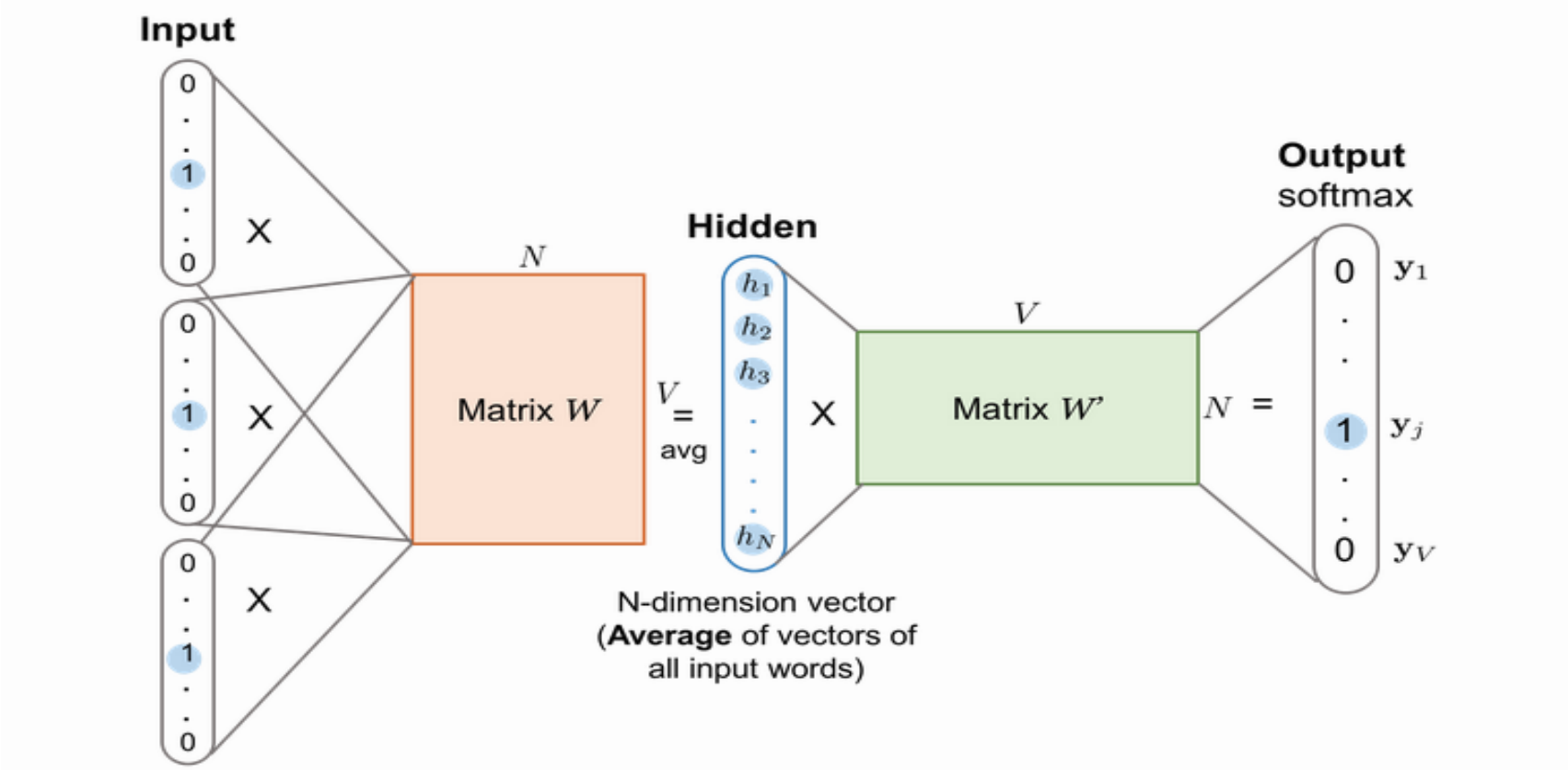


Fig. 2. The CBOW model. Word vectors of multiple context words are averaged to get a fixed-length vector as in the hidden layer. Other symbols have the same meanings as in Fig 1.