



# Natural Language Processing (NLP)

## Lecture 4

Word2Vec.. Part I: CBOW

# Word2Vec == Word to Vector

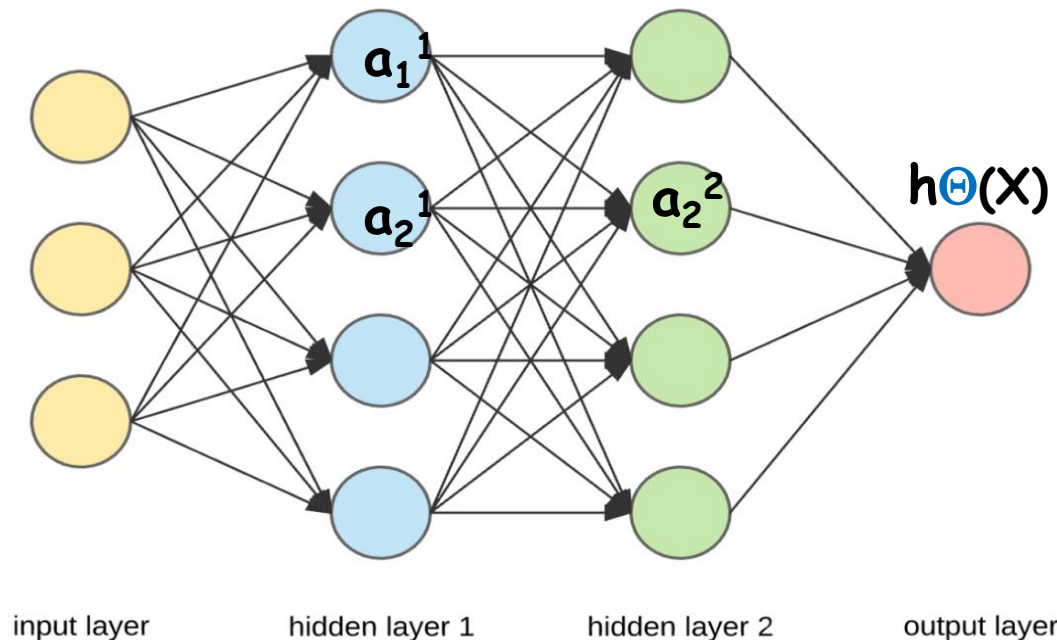
- Particular machine learning model
  - Extends/employs the word embedding concept
  - Uses 2 layers ANN model process text.
  - It obtain words associations based on word embeddings.
  - Derive the similarities between words to allow predicting the missing words.
- 
- **Two types:**
    1. **CBOW** : Continuous Bag Of Words  
Employs a group of occurred words to get a missing word
    2. **SG** : Skip Gram  
Employs a word to get a group of words.

# ANN in brief

- Consists of artificial neuron cells simulating the human brain.
- Contains one input layer to represent the input features.
- Contains N hidden layers as intermediate layers
- Contains one output layer with single/multi cells
- Cells are connected among each others with weights ( $\Theta_{jk}^i$ )

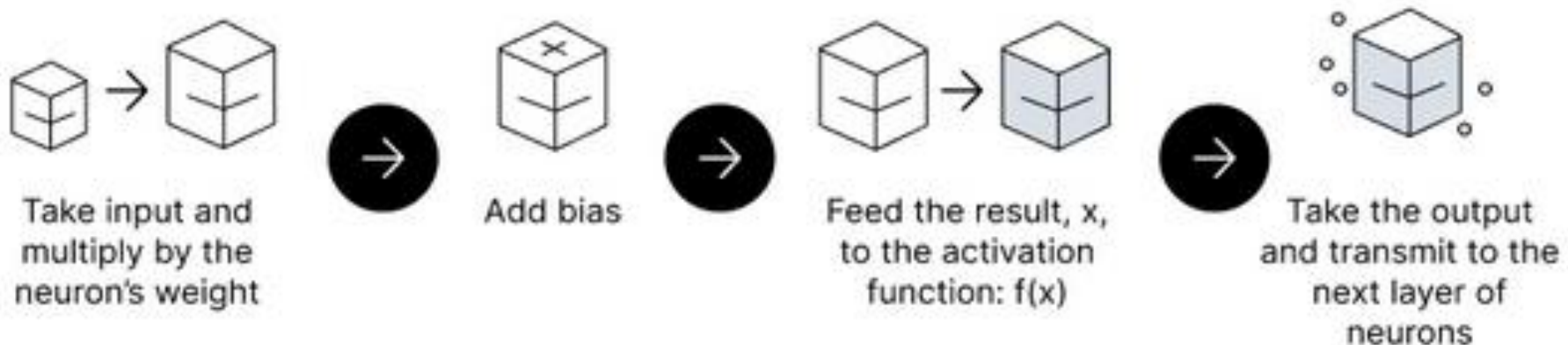
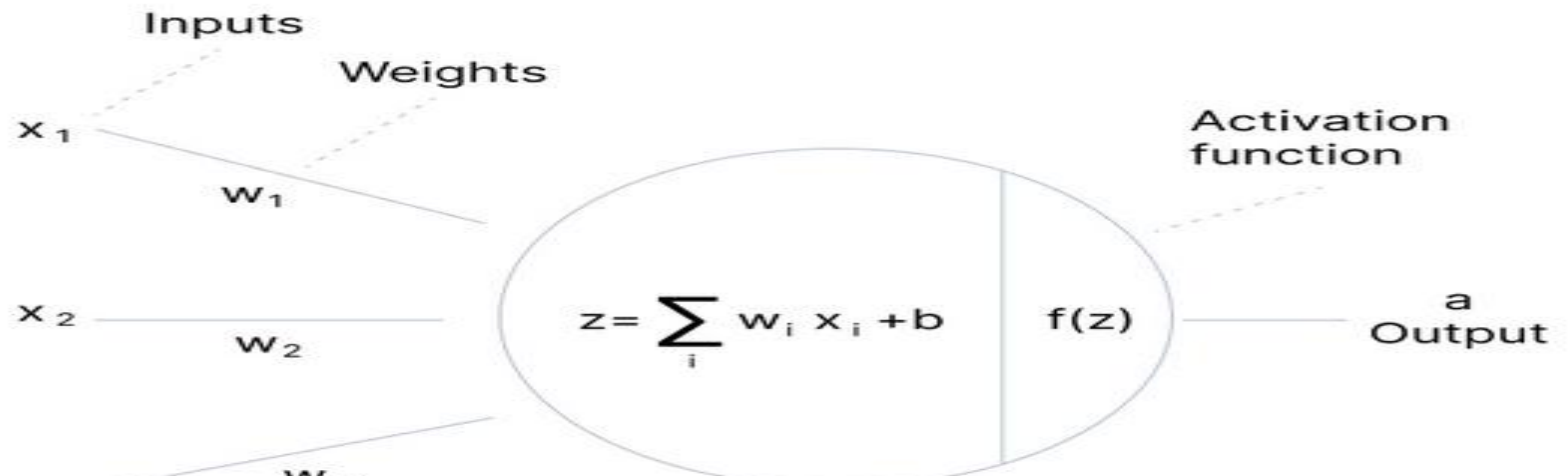
where weights are represented in an vector and i the layer num,  
jk connected cell numbers

- **Activation function ( $a_j^i$ )** :  
output of the cell  
where(i:layer num, j:cell num)



# Activation function

The primary role of the Activation Function is to transform the summed weighted input from the node into an output value to be fed to the next hidden layer or as output.



# Types of NN activation functions

- Binary Step Function
- Linear Activation Function
- Non-Linear Activation Functions :
  1. Sigmoid / Logistic
  2. Tanh Function (Hyperbolic Tangent)
  3. ReLU Function
  4. Leaky ReLU Function
  5. Parametric ReLU Function
  6. Exponential Linear Units (ELUs) Function
  7. Softmax Function
  8. Swish
  9. Gaussian Error Linear Unit (GELU)
  10. Scaled Exponential Linear Unit (SELU)

# ANN simple example

## Classification Example for IRIS data by DNN

input features (D=4)

Categories of Classification  
possibility output (K=3)

Speal.Length

Speal.Width

Petal.Length

Petal.Width

setosa

versicolor

virginica

W1, b1

W2, b2

hidden layer used to capture the potential patterns (H=6)

Input  
Layer

Hidden  
Layer

Output  
Layer

Clump Thickness

Uniformity of Cell Size

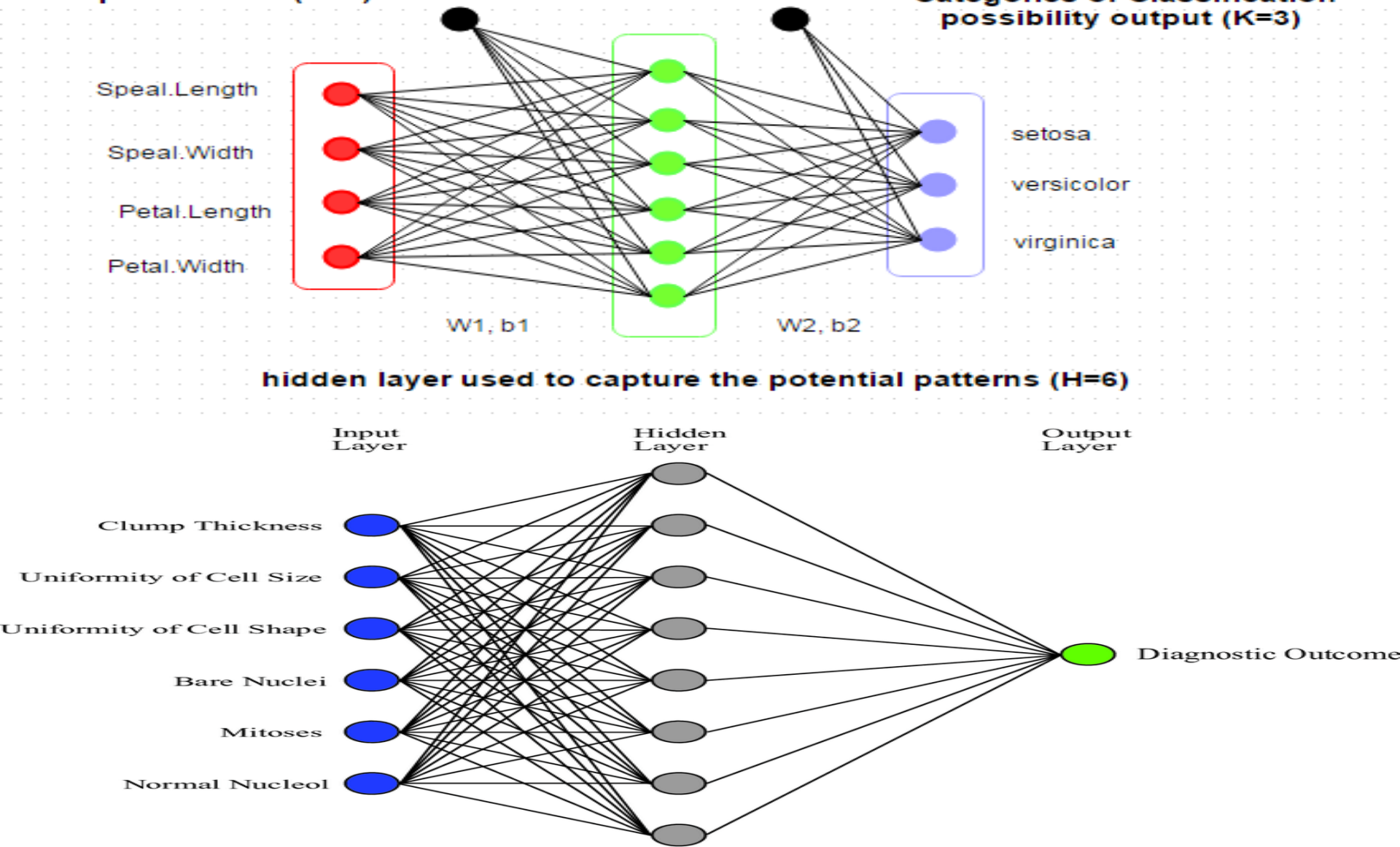
Uniformity of Cell Shape

Bare Nuclei

Mitoses

Normal Nucleol

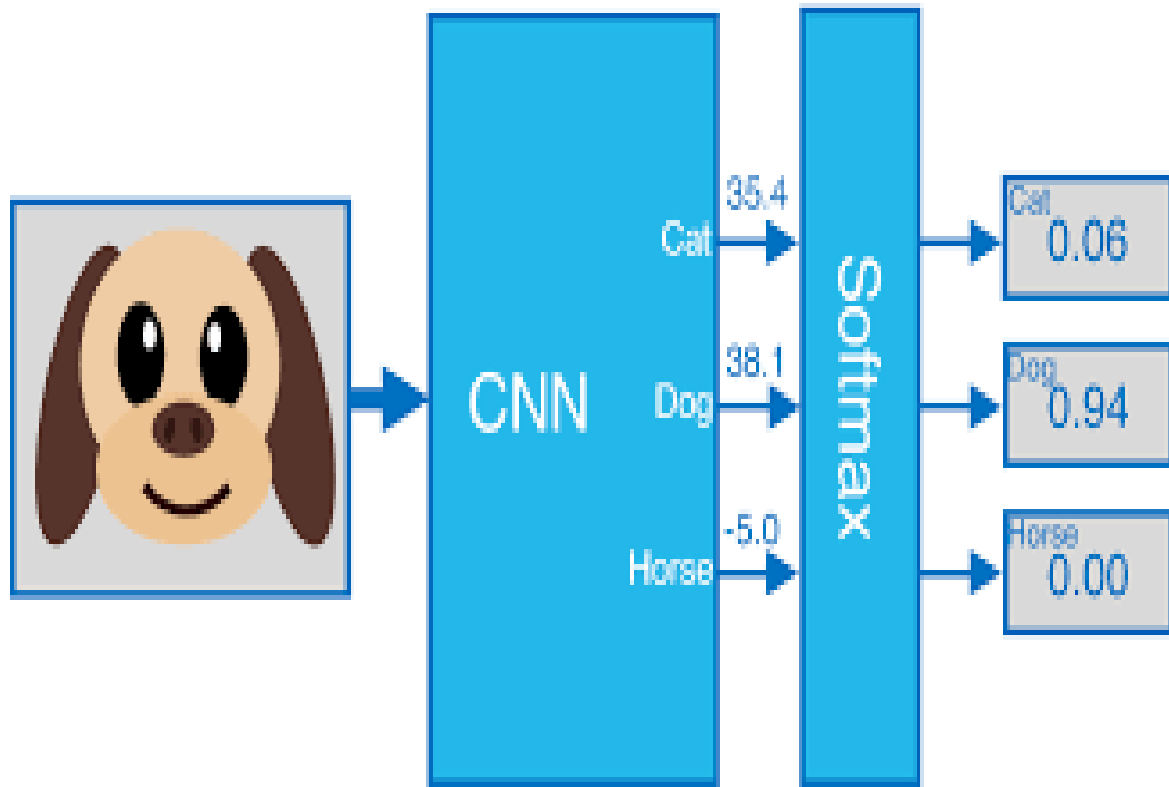
Diagnostic Outcome



# Softmax function

- An activation function for multi classification outputs of NN
- Produce equivalent probability to each class of the output classes
- Summation of probabilities at final layer = 100%
- Mathematical formula:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

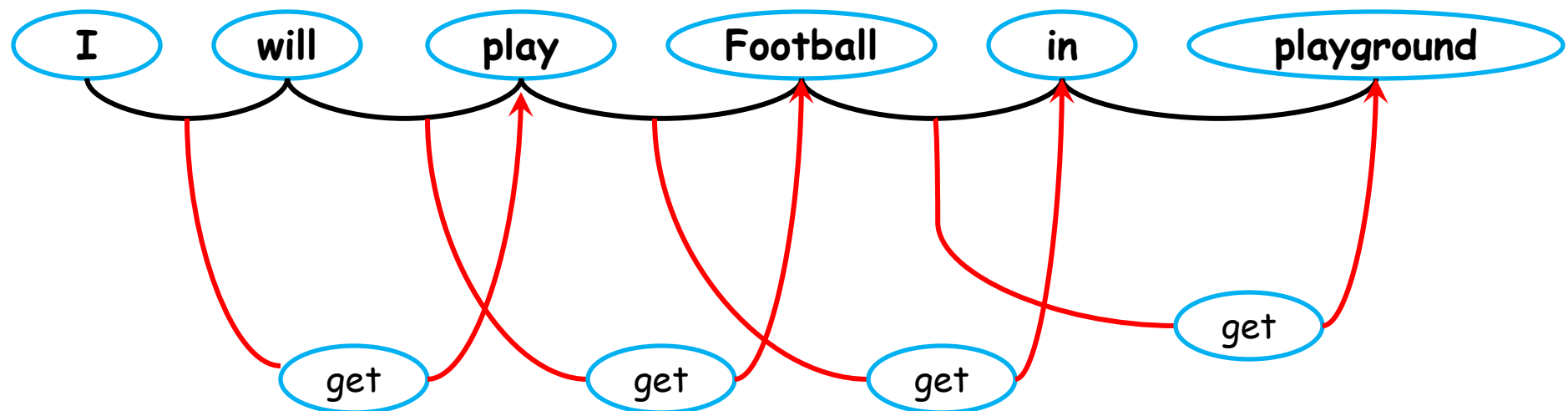


# CBOW

- Continuous Bag Of Words
- A Neural network in its core
- Based on <sup>1</sup>BOW (Bag Of Words) and <sup>2</sup>N Grams techniques.

1. **BOW**: calculate a matrix with its values represent the frequency of the word occurrences in the document
2. **N Gram**: based on using one or more of the existing words to predict the missing word.

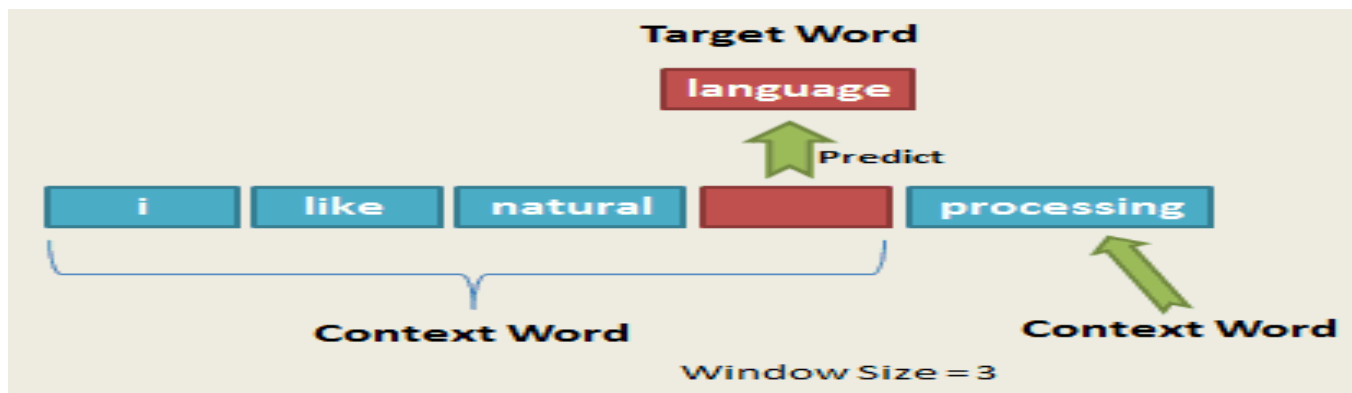
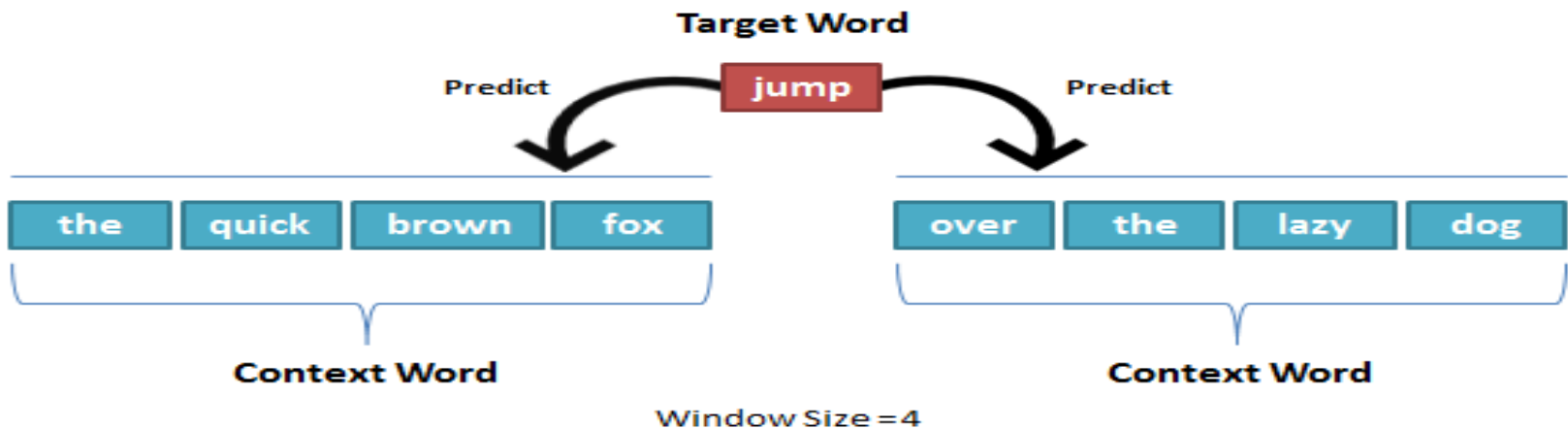
For example : Bi-Gram (**Text generation**)





# Window size

- In CBOW, not all the words in the document are employed in the processing.
- Window size = context size: number of words on both sides of the target word usually (5-10) words.



# CBOW Model

- No standard implementation for CBOW in Spacy & NLTK.
- CBOW model is efficient for the same data domain it were trained for (literature, medical, philosophy, ..).
- For the same data domain, CBOW model performance varies according to sub domains.

**For example:** different styles of writings in literature

# CBOW Model Flow

1. Data pre-processing (cleaning, lowering).
2. Tokenization & duplicates removal.
3. Determine **context size** (double no of input features on both sides of the target word ), **embedding size** (no of applied weights to CBOW NN ).
4. Preparing model learning samples → Build **word bags** (divide the text to sub-groups, each group consists of the words included in the context size as input features, and the target word as an output target/feature ).
5. Feed the model with the word bags groups to proceed in the training.
6. Proceed in NN training with applying Softmax activation function at the last layer.

# CBOW Applications

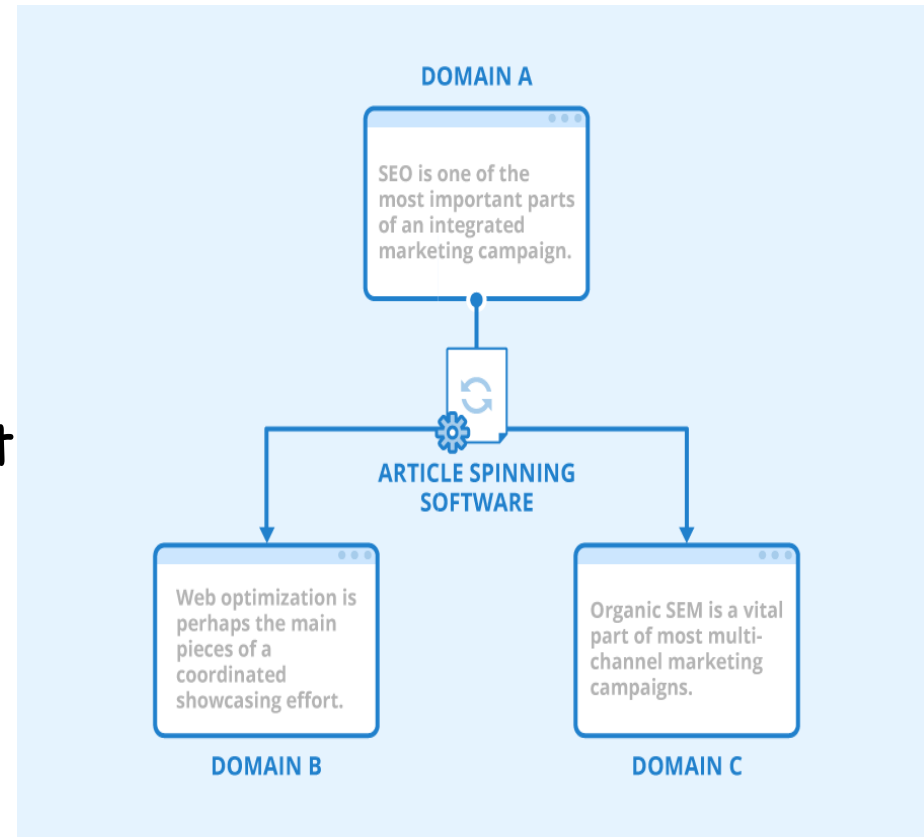
*For example*: Article spinning which is used to describe the rewriting of articles with keeping the original meaning, often using software, in order to make them seem unique to search engines.

## Why?

Web sites with copied articles are being under ranked from the search engines.

## How?

Remove some words from the original article and use CBOW to predict another substituting options to produce another document with different context but with Same/close meaning



THANK YOU ... 😊