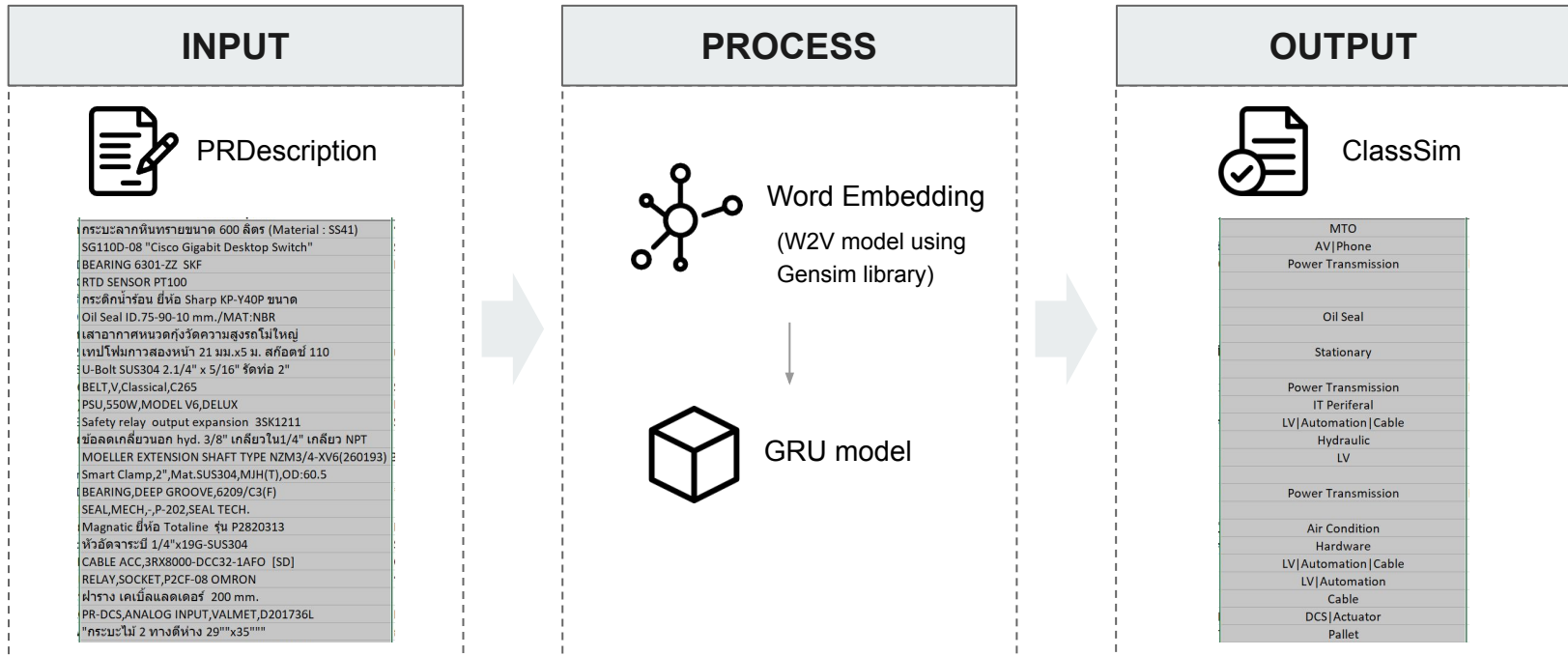




Purchasing Request Text Classification

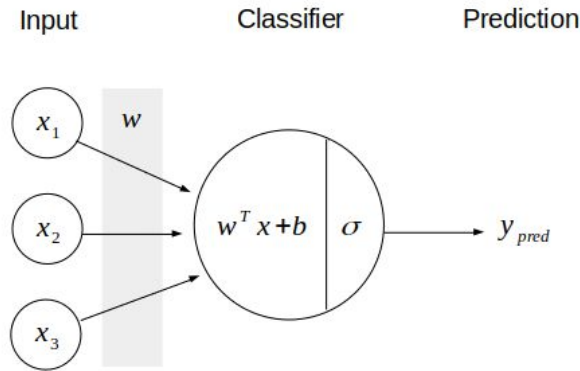
using Neural Network-based Algorithm

IPO Model Conceptual Framework

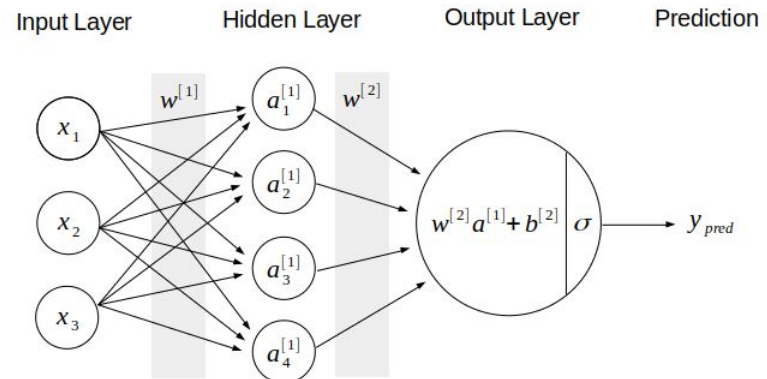


Neural Network Algorithm

Artificial neural network เป็นวิธีการที่มีพื้นฐานมาจากการเลียนแบบการทำงานของสมองมนุษย์ ประกอบด้วยนิวรอน (Neurons) ที่ใช้ในการเชื่อมโยงเพื่อแก้ปัญหาต่าง ๆ ซึ่งเป็นการเรียนรู้แบบหนึ่งของ Machine learning



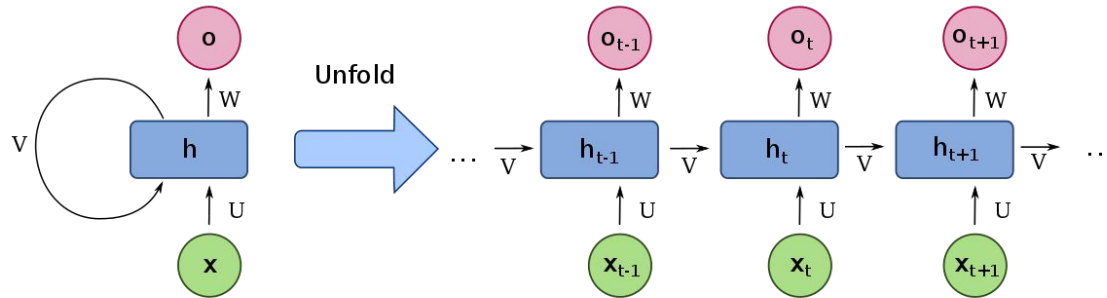
Machine learning algorithm ทั่วไป



Neural network algorithm

Recurrent Neural Network (RNN)

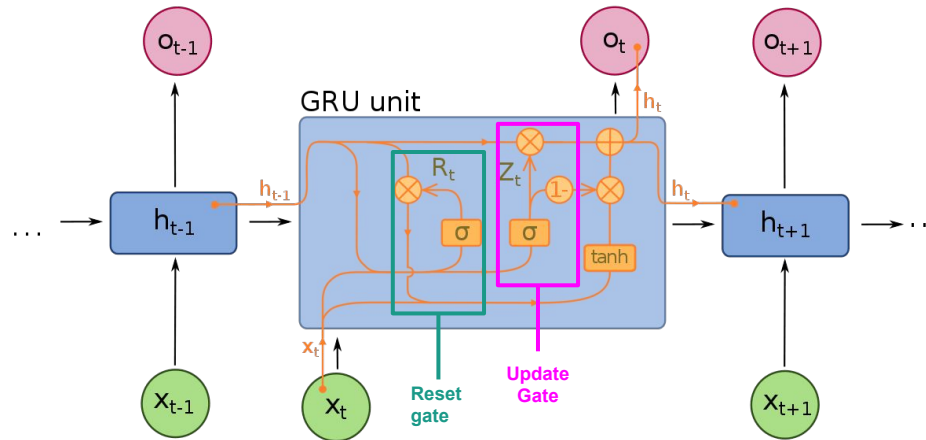
Recurrent Neural Network (RNN) คือ Artificial Neural Network แบบหนึ่งที่ย่อแบบมาแก้ปัญหาสำหรับงานที่ข้อมูลมีลำดับ Sequence โดยใช้หลักการ Feed สถานะภายในของโมเดล กลับมาเป็น Input ใหม่ คู่กับ Input ปกติ



A diagram for a one-unit recurrent neural network (RNN).

Gated Recurrent Unit (GRU)

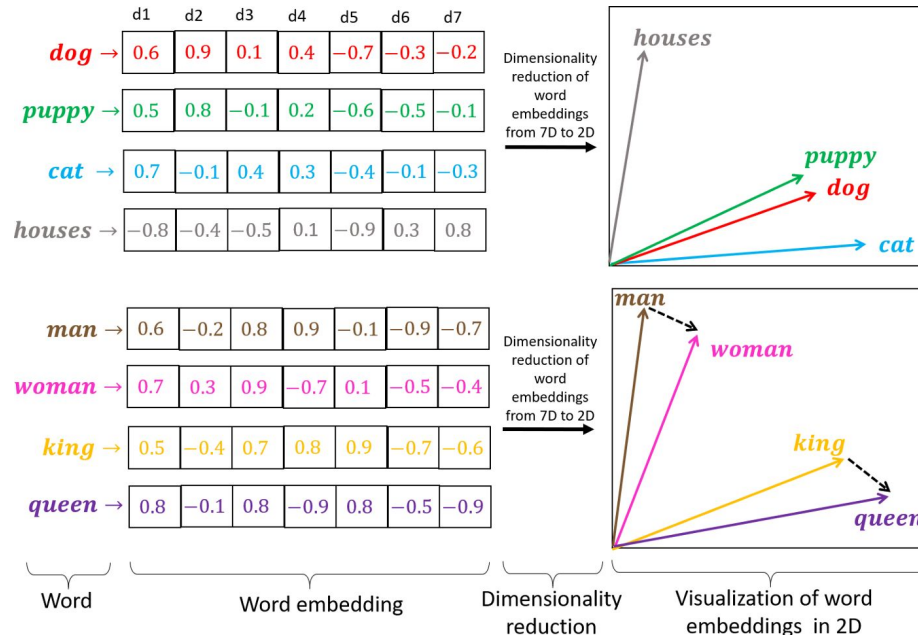
Gated Recurrent Units (GRU) เป็นกลไกปิดเปิดการอัปเดตสถานะภายใน Recurrent Neural Network



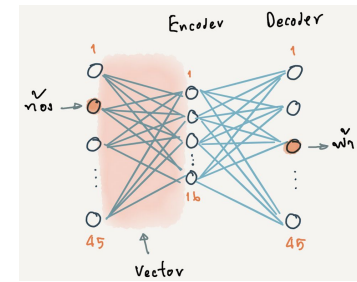
A diagram for a one-unit Gated Recurrent Unit (GRU).

Word Embedding

Word embeddings are vector representation of words learnt from context training.



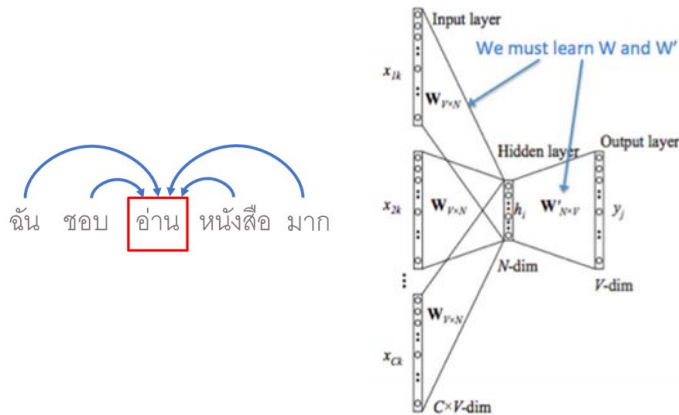
Word2Vec Model



การสร้าง Vector จากคำบริบทในประโยค โดยใช้เทคนิคทาง Neural Network แบบ Encoder-Decoder Model

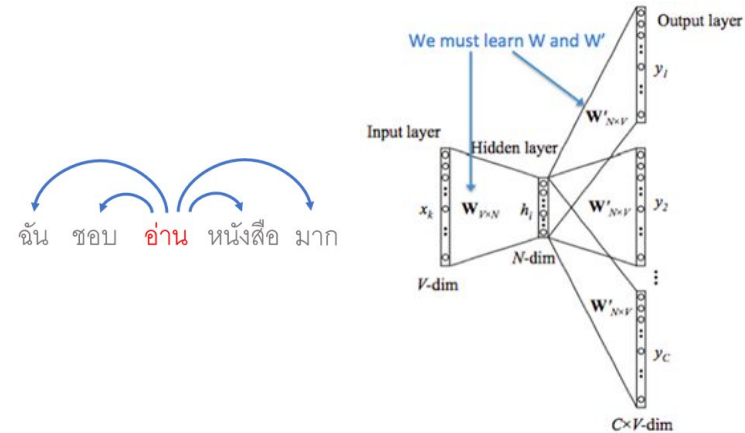
- Continuous Bag of Words (CBOW)**

คำคำหนึ่งนั้นอาจสามารถถูกทำนายได้จากบริบทของคำที่อยู่รอบข้าง



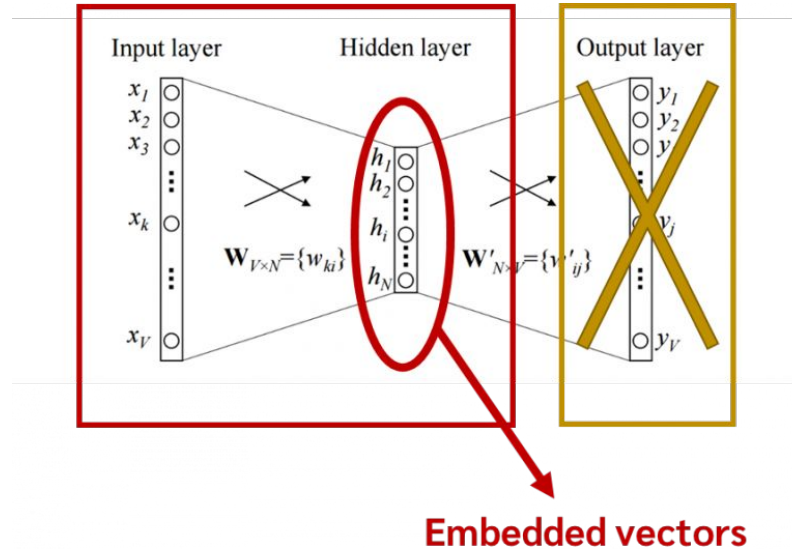
- Skip-gram**

คำที่อยู่รอบข้างของคำคำหนึ่งอาจสามารถถูกทำนายได้จากคำคำนั้น



Word2Vec Model

สิ่งที่เราต้องการจากโมเดล Word2Vec คือ ค่าความสัมพันธ์ อธิบายได้ว่า การเอาผลลัพธ์จาก hidden layer หลังจากใส่ word vector ของคำมาใช้แทนค่านั้นโดยตรง



Word2Vec Model using Gensim Library

Gensim เป็น Open source Python Library ในกลุ่ม NLP

```
from gensim.models import Word2Vec
```

```
model = Word2Vec(sentences, vector_size = 100, window = 5, min_count = 5, workers = 3, sg = 1)
```

- **vector_size** (*int, optional*) – Dimensionality of the word vectors.
- **window** (*int, optional*) – Maximum distance between the current and predicted word within a sentence.
- **min_count** (*int, optional*) – Ignores all words with total frequency lower than this.
- **workers** (*int, optional*) – Use these many worker threads to train the model (=faster training with multicore machines).
- **sg** (*{0, 1}, optional*) – Training algorithm: 1 for skip-gram; otherwise CBOW.

<https://radimrehurek.com/gensim/models/word2vec.html>



Python Scripts