**NLP Knowledge**

Here is all my collected technique for Natural Language Processing technique

**Bag of Words**

Un bag of words est une technique qui prends un document le découpe en mots, et pour chaque label donné il va compter le nombre de mots dans chacun de ses labels. Ensuite avec ce compte, on peut prendre une nouvelle phrases labelisé, créer un vector pour cette phrases de la taille du vocabulaire, et utiliser le one hot encoding sur chaque x du vector où est un mot de la phrase labélisé est dans le vocabulaire, ce qui nous donne un vector avec beaucoup de 0 et quelques 1, en utilisant ces vectors qu’on dit de type Sparse, on peut entrainer un réseau de neurones en passant ces énormes vecteurs dans l’input layer.

**Words Embedding**

**Word2Vec**

There are two

main training algorithms that can be used to learn the embedding from text; they are Continuous

Bag-of-Words (CBOW) and skip grams. We will not get into the algorithms other than to say

that they generally look at a window of words for each target word to provide context and in

turn meaning for words.

Glove lui se concentre sur les caractères autour du mot

Voir GaussianEmbedding → Word2Gaussian

- Un mot est reconnu en fonction des mots qui l’entoure

- Prendre un mot et compter sa fréquence dans X Documents de différents types Y,

si la fréquence d’apparition du mot dans chaque document se rapproche de la fréquence d’apparation dans un autre mot alors on peut déduire un rapprochement

**End-to-end Memory Network (MemN2N)**

Good pour : Q&A

Le but est de pouvoir prendre des phrases ordonné dans le temps, et en créant des embedding vectors pour chaque mot et en ajoutant un vector de (time embedding) pour chaque phrase en fonction de leurs ordre d’apparition dans le texte, on peut alors définir répondre a des questions simples sur le texte.