

Projet Intelligence Artificielle

Algorithme de prédiction des mots

Word2Vec

BRUNET Nathan

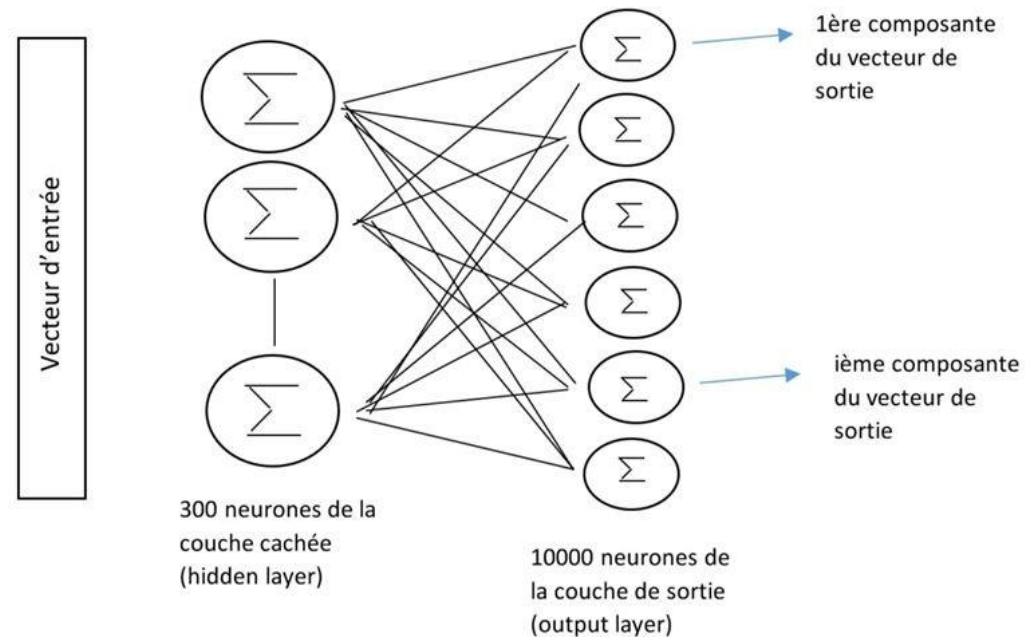
MICHOT Vincent

Sommaire :

- Choix de la démarche
- Mise en œuvre

Choix de la démarche :

- ▶ Création d'un dictionnaire
- ▶ Vectorisation des mots (Word2Int)
- ▶ Création du réseau de neurones :
 - première couche (cachée - « hidden layer »)
 - couche finale (utilisation de SoftMax)
- ▶ Renvoi du mot ayant la plus grande probabilité de suivre le mot d'entrée



Mise en œuvre :

- ▶ Indexation des mots du dictionnaire - vectorisation
- ▶ Choix d'une fenêtre de taille 2 autour du mot d'entrée du texte brut initial
- ▶ Choix d'une couche cachée (« hidden layer ») de 2 neurones
- ▶ Détermination du poids des mots du dictionnaire par la seconde couche
- ▶ Obtention en sortie de coordonnées dans un plan 2D pour les mots du dictionnaire
- ▶ Calcul de la distance du mot d'entrée à chacun des mots du dictionnaire
- ▶ Renvoi du voisin suivant le mot d'entrée avec la plus grande probabilité (situé au plus proche de ce dernier dans la cartographie).