idl_17 : Classification de documents

Idée : attribuer la bonne catégorie à un document.

❓ Quels sont les exemples que vous connaissez ?

- méthode non supervisée
  On cherche ici des stratégies pour se passer de données annotées.

Exemple de méthode non supervisée :

Quel est le thème principal de cette phrase ?

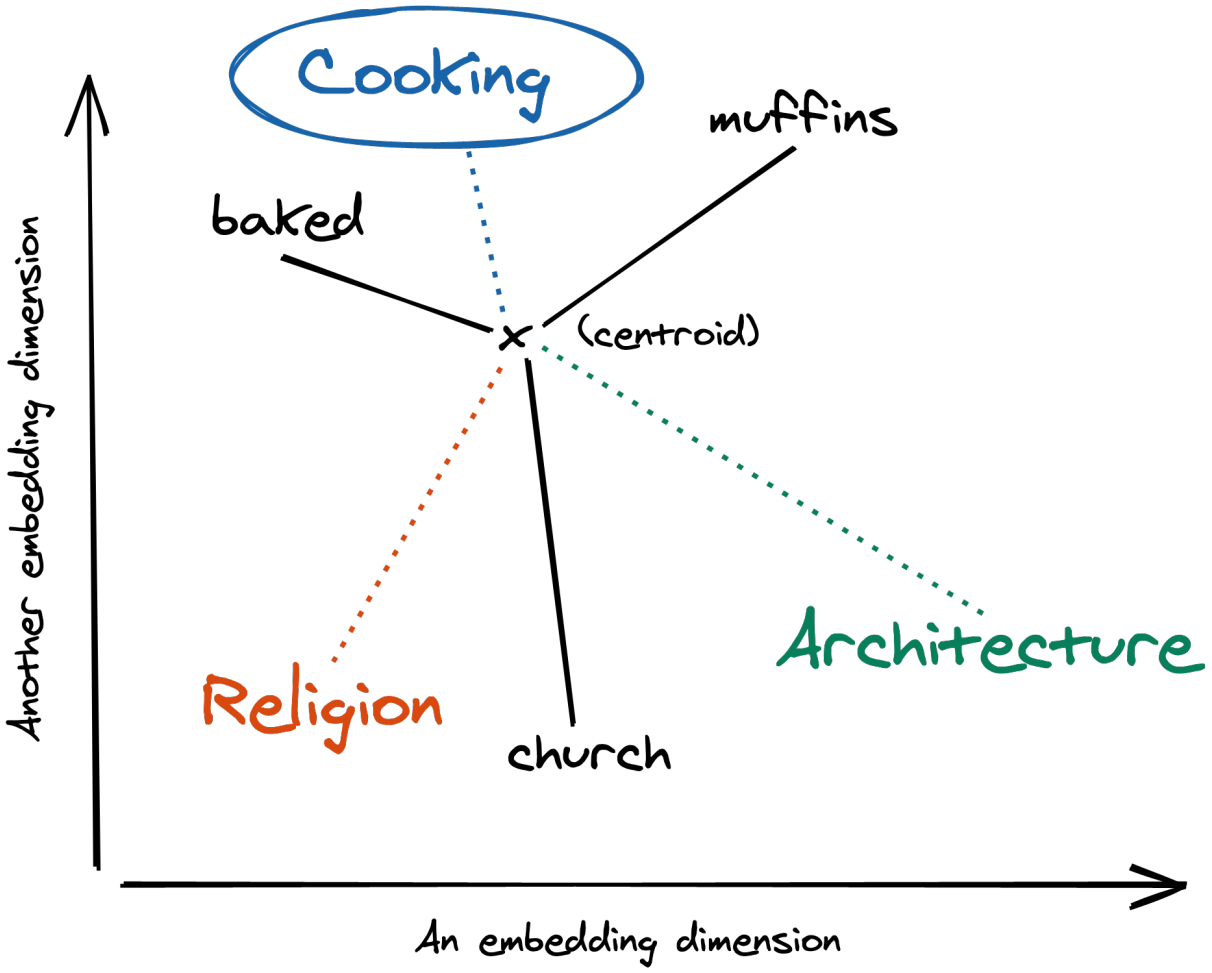
- méthode supervisée : demande beaucoup de données labélisées (TP de demain)
  On cherche ici à trouver une fonction qui à partir de données d'entraînements et de leur étiquette, peut deviner l'étiquette (la classe) d'un nouveau document.

```
I brought some muffins to church, I baked them myself.
```

- cooking ?
- religion ?
- architecture ?

La méthode :

1. on regarde les vecteurs des mots qui composent le document, et on calcule le centroïde de ce vecteur.
2. on calcule la distance entre ce centroïde et les vecteurs des classes elles-mêmes pour trouver le plus proche voisin
3. on choisit la classe qui a la plus faible distance



Exemples de méthodes supervisées

Discussions