

Questions pour Discussion (LAMBARAA Abdellah)

Différence entre distance et similarité :

→ La distance mesure la **différence**, la similarité mesure la **ressemblance**.

Pourquoi normaliser ?

→ Pour éviter qu'une variable à grande échelle (ex : revenu) domine les autres.

Distance la plus sensible aux valeurs extrêmes :

→ La **distance euclidienne**, car elle élève les différences au carré.

Quand préférer la similarité cosinus :

→ Quand on veut comparer la **forme** des profils (direction du vecteur) plutôt que leur **magnitude** — par exemple dans les données de texte ou les comportements utilisateurs.

Conclusion courte

- La **distance euclidienne** augmente quand deux clients sont **différents**.
- La **similarité cosinus** augmente (proche de 1) quand deux clients sont **semblables**.
- Donc, plus la distance est **petite**, plus la similarité est **grande**.
- Les deux mesures traduisent la proximité, mais en sens inverse.