

MATEMÁTICAS Y DATOS IX... in a nutshell

La distancia de Wasserstein entre distribuciones



Bien entendida una base de datos es un muestreo sobre un conjunto de variables aleatorias las cuales en un principio podrían seguir leyes de probabilidad distintas. Calcular la distancia entre dos distribuciones de probabilidad es una pregunta natural.

En 1969 el matemático ruso Leonid Vaserstein propuso una métrica para calcular la distancia entre dos distribuciones inspirada en un problema planteado en 1781 por Gaspard Monge.

Si pensamos en una distribución como una montaña, transportarla de un lugar a otro para convertirla en otra montaña requiere de tanta energía como la cantidad de arena multiplicada por la distancia entre las dos.



La distancia de Wasserstein es particularmente útil para distribuciones en muchas dimensiones, pensemos por ejemplo en bases de datos de imágenes o de textos.