

Redes neuronales profundas generativas GAN

Son una técnica de aprendizaje profundo que se utiliza para generar datos sintéticos similares a los datos de ~~entrenamiento~~ entrada, y estos pueden ser utilizados para agrupar los datos en diferentes clusters.

Las GAN autoencodoras es una variante y consta de dos redes neuronales, un generador y un discriminador.

El generador se utiliza para generar datos sintéticos que se parecen a los datos de entrada, y el discriminador evalúa que tan buenos son los datos generados comparados con los datos reales, luego el autoencoder se utiliza para comprimir los datos de entrada en un espacio de menor dimensión, que se pueden utilizar para hacer clustering.

También otra variante es la GAN de clustering, donde hace clustering directamente utilizando la red. El generador se utiliza para generar representaciones ~~reales~~ de datos sintéticos que se ajustan a los clusters identificados en los datos de entrada. El discriminador, toma la muestra de datos ya sea real o sintética y determina si es real o sintética.

Durante el entrenamiento, el GAN trata de maximizar la probabilidad de que el discriminador no pueda distinguir entre los datos reales generados por el generador y los datos reales durante el entrenamiento. El generador trata de maximizar la probabilidad de que el discriminador se equivoque y clasifique los datos sintéticos como reales.

Una vez entrenada la red, se utiliza la aprendida por el generador para hacer clustering, utilizando un algoritmo como K-means o DBSCAN, en el espacio aprendido por la red.

Estos datos se utilizan en situaciones que los datos son complejos y no se agotan fácilmente a una representación lineal. En estos casos, el GAN puede aprender una representación no lineal de los datos que sea más adecuada para hacer clustering.