Machine Learning para preservar la anonimidad

Normalmente las técnicas de machine learnign deben manegar volúmenes de data enormes, lo que requiere muchos recursos de computación, el uso de nodos independientes no es apto debido a límites de memoria o de tiempo. Algunos de los frameworks más usados para machine learning están optimizados para usar GPUs. Existen muchos sistemas distribuidos de computación capaces de realizar entrenamiento en múltiples nodos de forma coordinada. Las aproximaciones convencionales requieren de una plataforma centralziada que recoge los datos y los distribuye entre los nodos.

Técnicas de preservación de privacidad en Machine Learning

- **Computación Multi-Grupo Segura**: Una función puede ser computada colectivamente por varios grupos sin mostrar sus propios datos. Cada grupo debe intercambiar sus salidas con otros en secreto para que puedan ser agregados. Es una de las formas más seguras de entrenar modelos de machine learnign. Por desgracia no es bueno apra entrenar modelos muy complejos o a gran escala.
- **Cifrado homomorfico**: Puede ser aplciada directamente para cifrar datos que luego son transferidos a un servidor central. Cualquier modelo puede, en teoría, ser entrenado con estos datos cifrados, obteniéndose un modelo cifrado que puede ser enviado a los clientes. El server no es capaz de descifrar ni los datos ni el modelo. Los clientes tienen acceso a sus propios datos y al modelo descifrado una vez entrenado. Este modelo tampoco es adecuado para larga escala.

Aprendizaje Federado