el col·loqui

25 d'octubre de 2023

Centre de Recerca Matemàtica
Aula A1
A les 12.00h

Josep Domingo-Ferrer

Universitat Rovira i Virgili

Privadesa en aprenentatge automàtic centralitzat i descentralitzat

L'aprenentatge automàtic (ML de les sigles en anglès) és vulnerable a atacs a la seguretat i a la privadesa. Mentre que els atacs a la seguretat volen impedir la convergència del model o forçar la convergència a un model erroni, els atacs a la privadesa miren d'obtenir les dades que s'han fet servir per entrenar el model. Aquesta xerrada se centrarà en els atacs a la privadesa. Després de passar-hi revista, examinaré l'ús de la privadesa diferencial (DP de les sigles en anglès) com a metodologia de protecció, tant en ML centralitzat com descentralitzat (aprenentatge federat). Mostraré que les implementacions del ML basades en la DP no forneixen les garanties de privadesa "ex ante" de la DP. El que forneixen és essencialment addició de soroll semblant a l'aproximació tradicional del secret estadístic. El nivell real de privadesa assolit ha d'avaluar-se "ex post", cosa que es fa rarament. Presentaré resultats empírics que mostren que les tècniques estàndard per evitar el sobreaprenentatge (overfitting) en ML donen un compromís millor entre utilitat/privadesa/eficiència que la DP.

UAB
Universitat Autònoma de Barcelona

Departament de Matemàtiques