Projet : Attaque et Défense dans l'apprentissage fédéré

Chuan Xu

chuan.xu@univ-cotedazur.fr

2025

Projet : Attaque et Défense dans l'apprentissage fédéré

Date de rendu: 17/04/2025

Projet : Attaque et Défense dans l'apprentissage fédéré

Date de rendu: 17/04/2025

Demo Code: `https:`

`//gitlab.inria.fr/chxu/fl_miage_ia_2025/-/tree/main/Projet?ref_type=heads`

- Projet en Binôme
- Tâche: Écrivez les attaques et tester les défenses dans l'apprentissage fédéré
- Jeu de données: CIFAR 10

Attaques

- Inversion d'étiquettes (Binôme I)
  1. Inverser étiquette ($0 -> 1, 1 -> 2 ...$)
  2. L'attaquant exécute cette attaque aléatoire.
- Altération du modèle (Binôme II)
  1. L'asente de gradient
  2. L'attaquant exécute cette attaque aléatoire.

Défenses (Binôme I et II)

- Médiane par coordonnées (FedMedian)
- Moyenne tronquée$_f$ (FedTrimmedAvg)
  f indique le nombre d'attaquant. Dans FedTrimmedAvg, *beta* indique la proportion d'attaqaunt.

Évaluation

- 15% Code
  (avec les commentaires et le retour de fichiers qui enregistrer les statistiques) (codes pour les plots)
- 85% Rapport

Rapport

- Introduction de l'apprentissage fédéré et FedAvg (1 point)
- Introduction les deux attaques et les deux défenses (1 point)
- Résultats (18 points)
  1. Jeu de données, le modèle, scénario tester, la distribution de jeu de données (1 point)

Rapport

- Résultats (18 points)
  1. Jeu de données, le modèle, scénario tester, la distribution de jeu de données (1 point)
  2. Attaques (deux figures chacun, 8 points)
     1. Effet de nombre de attaquants dans le cas iid (2 points)
     2. Effet de nombre de attaquants dans le cas non-iid (2 points)
     3. Analyses de chaque figure et Comparer les deux figures (3 points)
     4. Si binôme, il faut aussi compare les deux attaques (1 point)

     Attention : chaque configuration, attaque 5 fois et montrer le résultat moyenne et aussi la déviation !

Rapport

- Résultats (18 points)
  1. Jeu de données, le modèle, scénario tester, la distribution de jeu de données (1 point)
  2. Attaques (deux figures chacun, 8 points)
  3. Défenses (trois figures chacun, 9 points)
     1. Effet de nombre de attaquants avec *Médiane par coordonnées* dans le cas iid (2 points)
     2. Effet de nombre de attaquants avec *Moyenne tronquée*$_f$ dans le cas iid (2 points)
     3. Analyses de deux figures et comparer les performances de deux défenses (2 points)

Rapport

- Résultats (18 points)
  1. Jeu de données, le modèle, scénario tester, la distribution de jeu de données (1 point)
  2. Attaques (deux figures chacun, 8 points)
  3. Défenses (trois figures chacun, 9 points)
     1. Effet de nombre de attaquants avec *Médiane par coordonnées* dans le cas iid (2 points)
     2. Effet de nombre de attaquants avec *Moyenne tronquée*$_f$ dans le cas iid (2 points)
     3. Analyses de deux figures et comparer les performances de deux défenses (2 points)
     4. Effet de nombre de attaquants avec *Médiane par coordonnées* dans le cas non-iid (2 points)
     5. Analyses et comparer avec le cas iid (1 point)

Rendu

Rendu sur Moodle (code zippé et un rapport pdf)