

Biais et éthique en apprentissage statistique

Nicolas Leroy, Ema Cerezo, Axel de Montgolfier

2023-10-16

Enjeux

- ▶ Qu'est ce qu'un biais ?
- ▶ Quelles sont les définitions d'un système éthique ?
- ▶ Quelles méthodes pour suivre de telles définitions ?

Cas d'application

- ▶ COMPAS : Estimateur de la capacité d'un criminel à récidiver adopté en 2016 par l'état du Wisconsin.
- ▶ Gender Shades project : Tentative en 2018 de rendre plus éthique le système de reconnaissance faciale par la repondération d'échantillon.

Deux définitions d'un modèle éthique

Equalized Odds (Égalité des erreurs) : Veille à ce que les groupes bénéficient de taux de faux positifs et de vrais positifs égaux, indépendamment de la variable sensible.

$$\mathbb{P}(\hat{Y} = 1|S = 1, Y) = \mathbb{P}(\hat{Y} = 1|S = 0, Y)$$

Demographic Parity (Parité démographique) : Garantit que les individus ont des chances égales d'avoir un résultat favorable quel que soit leur groupe.

$$\mathbb{P}(\hat{Y} = y|S = 0) = \mathbb{P}(\hat{Y} = y|S = 1)$$

Catégories de biais

- ▶ Biais de collection de données
- ▶ Biais de modélisation
- ▶ Biais d'utilisation

Comment remédier aux discriminations

- ▶ Correction de données (pre-processing)
- ▶ Modèle résilient (in-processing)
- ▶ Correction de résultat (post-processing)

Conclusion

- ▶ Le champs recherche autour de l'éthique en apprentissage statistique se développe
- ▶ Il existe des solutions libre qui permettent de mesurer et d'améliorer le respect de l'éthique de nos modèle