Differentially Private Measures of Statistical Heterogeneity

Mary Scott, Graham Cormode and Carsten Maple

Outcomes

Explore existing and emerging research in **Federated Machine Learning** systems:

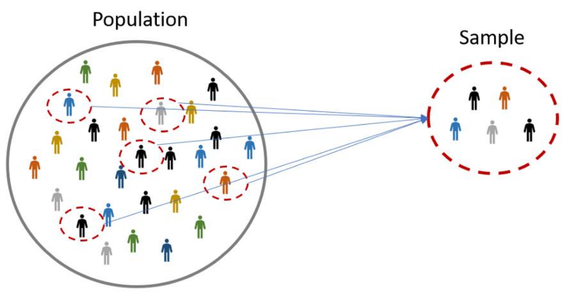
- which tackle problem of **statistical heterogeneity**,
- while incorporating **differential privacy**.

Two examples of FML systems

- *Federated Learning* – an alternative to traditional centralised ML techniques.
- *Federated Analytics* – can measure, analyse and improve FL models.

Two examples of non-i.i.d.

- *Skewness* – data deviates from a normal distribution through a shift to one side.
- *Statistical (Data) Heterogeneity* – asymmetry between different distributions of different subpopulations within the dataset.



Want to **minimise difference between observed and general trends** of entire dataset.

Differential privacy

Leading candidate to provide privacy protection.

- Cannot isolate a particular user's data from dataset
- Cannot determine **whether particular user is in dataset**
- Mathematically strong and has provable guarantees

Helps to keep **political and medical information** private.

Contact

Mary Scott

Mary.P.Scott
@warwick.ac.uk