Collecting data in Med Practice is low vol.

↳ because little data, anomalies more freq./relevant.
→ volume important but also variety

- Currently share across hospital
  ↳ EHR, handwritten notes useful but contains private data.

PROBLEMS:

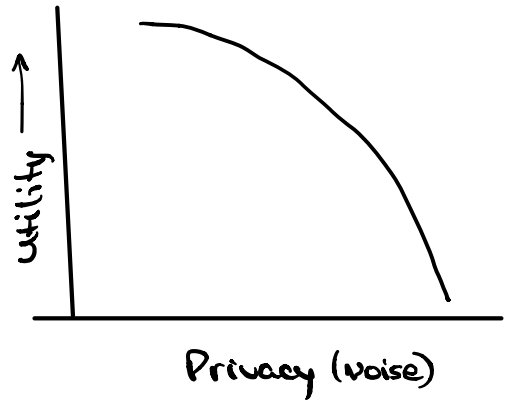
① Patient not responsible for privacy

② Anonymized data (masking)
- Can correlate different data sets and reverse engineer anonymized data

DIFFERENTIAL PRIVACY

anything learnable via Statistical Querying.

inject noise into the data, and account for it.
for privacy of the submitter not the aggregator.

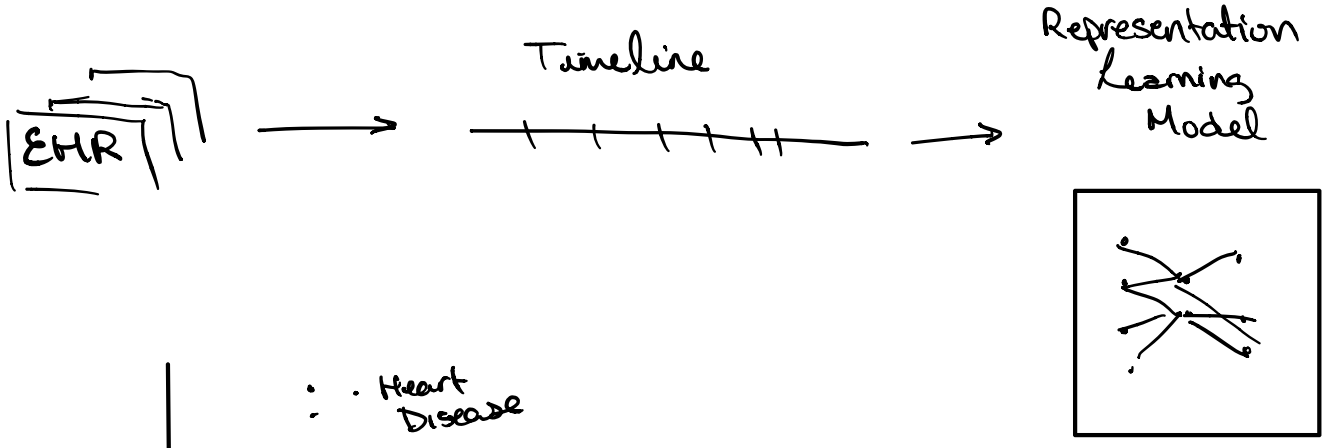


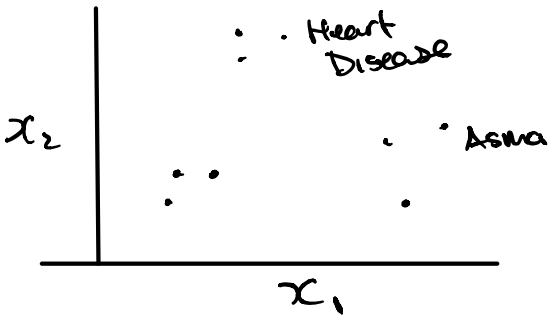
HOW PRIVACY & DL WORK

- DL on unstructured data
  ↳ hard to find identifiers?
  ↳ does model leak info about data/similar to overfitting

Maintaing Privacy
Apply differential privacy or APIs—distributed, build node wherever.

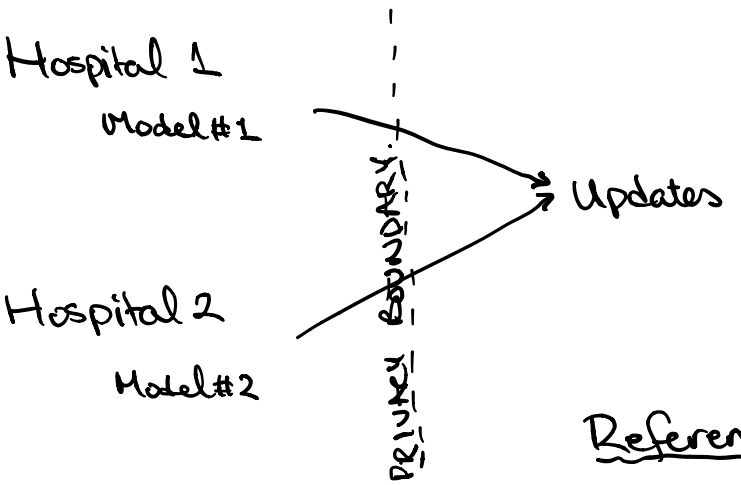
How





Can inject more noise with more data

"A federated learning technique."



References

I. Foundations & Trends in Theoretical C.S.
Dwork C. and Roth.

II. Deep Learning with.
Abdi?