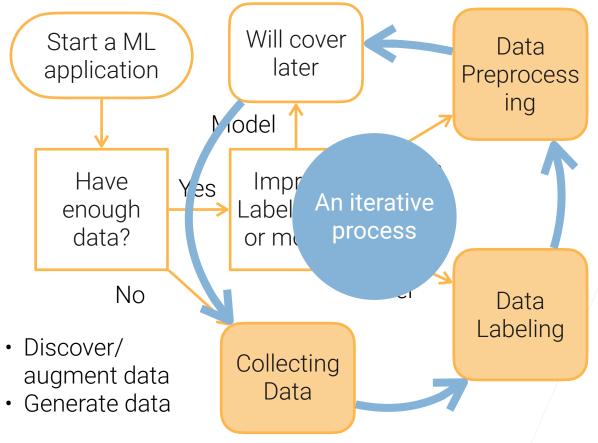# 2.5 Data Part Summary

Qingqing Huang, Mu Li, Alex Smola

https://c.d2l.ai/stanford-cs329p

Flow chart

- Exploratory data analysis
- Data cleaning
- Data transformation
- Feature engineering

- Semi-supervised learning
- Label via crowdsourcing
- Weak supervision

Start a ML application

Will cover later

Data Preprocessing

Have enough data?

Yes

Model

Impr Label or m

An iterative process

No

Data Labeling

- Discover/ augment data
- Generate data

Collecting Data