Data-Driven Systems

Data Modalities
- Tabular
- Textual
- graph
- Multimodal (image, ...)
- ...

Data → Alg/model → For a Task

Data: How to obtain the Data.

Alg/model: How to Formalize the Problem

For a Task:
- Student Success Predic
- Recidivism
- Clustering
- ...

Task/objective examples
- Utility maximization
- Classification/clustering
- Resource Allocation
- ...

Performance metrics:

Alg. : Time/Space Complexity

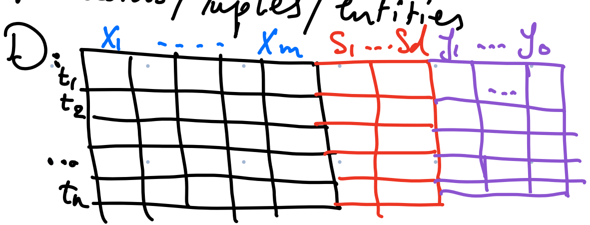
ML models: Accuracy, F1, ...

New metrics:

Fairness, Robustness, Stability, Transparency, Explainability, Privacy, ...

Notations:

DataSet Model: A Set/Collection of Records/Tuples/entities



Observations/Features attributes

$X = \{x_1 \dots x_m\}$

↳ Observable during the inference time

↳ Used for generating the outcome

Target Variable(s)/Label(s): $y_1 \dots y_o$

↳ available during the Training Time but Not the inference Time

Predic y as a function of X

$y \sim \underbrace{g(h_\theta(x))}$

↳ Simplified as $h_\theta(x)$

Training obj.

Find θ s.t.

$$\min_{t \in D} Loss(h_\theta(x), y)$$

$S = \{S_1, \dots, S_d\}$

↳ a Collection of Sensitive attributes used for identifying the Protected groups or demographic groups

e.g.

S: race, sex, age-group, ...

G: {black, white, female, male, ...}

Intersectional groups:

defined over the Cartesian of Sensitive attributes

↳ {Black Female, ...}

we may/may not have access to the sensitive attributes at the inference time.

q_1: How to define the groups?

q_2: which groups to consider?

q_3: How to obtain the group info.

- Feature Definition/Discretization
  - categorizing the scores of students {A, ..., F}
  - where to draw the Boundary.

Data Collection:
- Can Introduce Error/bias.

e.g., Survey:

1: Are you interested in responding to surveys

☐ yes ☐ No

Collected Data
- yes
- No
- Sampling Bias

- Labeling Bias:
  - what label should be used
    - GPA
    - 2-year success of Alumni
    - Graduation Record
  - How to ensure bias of labler is not included?
    - StereoTypes

- Proxy Attributes
  - Target Variable Dillema
    - we usually do not have access or cannot measure the Target variable
    - e.g., - Target Var.: - who commited crime.
      - which student was successful
  - Resolution: Use measured features that "represent" the target variable
    - Called PROXY ATTRIBUTES
    - e.g., - Arrest Record $\xrightarrow{\text{Proxy for}}$ Crime Record
      - GPA $\xrightarrow{\text{Proxy for}}$ Student Success
  - Proxies become problematic when they are proxies for sensitive attributes as well.

- In general, any of the features can play as proxies for sensitive attributes

e.g., Salary $\xrightarrow{\text{Proxy for}}$ gender

Proxy attributes are the reason why fairness-Through-Unawareness* is not possible.

*: Removing / do not recording the grouping information & do not using it for decision making.

e.g., A loan risk predictor that puts a high weight on income will have gender-bias.

Disparate Treatment.

Explicitly using group information for Decision making

→ group membership is used as the input to the model

→ different Treatment of people based on their group membership

e.g., having different grading rubrics for male/female students.

Masking:

Purposefully, generalizing or masking some features to draw Biased Conclusions that are Justified by data!

Aka.; digital Red lining

- e.g.,
  - Selecting features that Prevent Certain groups from receiving Loan!
  - Cherry-Picking Ranking mechanisms that Put Certain Universities in high rank Positions

Disparate Impact (Implicit)

when the decision outcomes are different for different groups.

$$h_\theta(X) \perp\!\!\!\perp S$$

e.g., Recidivism Scores

$\{x_1, x_2, \ldots, x_m\}$

$\checkmark x_i$ Correlation(x_i, y) $\uparrow$ x_i should be a proxy for y

Correlation(x_i, S) $\downarrow$ x_i " not " " " S

S not being Predictable by $\{x_i\}_1^n$

y to be " " $\{x_i\}_1^n$

In Practice

Min Corr(X, S)

Max Corr(X, y)