CS 260 Notes for 10/09/2024

I. Algorithm runtime / Big-O notation
   A. Largely dependent on the # of data points and data structures in use
   B. Constant time operations (`O(1)`): accessing values in an array, dictionary, etc.
   C. Iterating through `n` items: `O(n)`
   D. Sequential steps: add their runtime
   E. Nested operations: multiply their runtime
   F. Big-O notation: ignore scalars and very small terms

Example 1:

```
 l_m = []
 l_n = []
for items in l_m:
  do s+h // O(m)
for items in l_n:
   do s+h // O(n) runtime
```

Example 2:

```
  for m:
     for n:
        do s+h // O(m*n) runtime
```

II. Algorithm fairness
   A. Are algorithms fair by default? Unfairness manifests in a variety of ways:
      1. **Sample size disparity** – algorithms trained on a dataset with a majority / minority distinction perform worse for the minority by default
      2. ML is not fair by default, even though it relies on the neutrality of math
      3. If training data reflects social biases, the algorithm will likely incorporate them. Consider the example of **machine translation trained on a limited dataset**
   B. Challenges: Algorithms can…
      1. Inherit the prejudices of of their designers
      2. Reflect cultural biases
      3. Make it hard to identity biases
      4. Deny historically disadvantaged groups from full participation
   C. How can we tell if an algorithm is biased?

Example:

`Consider` a dataset `D` with attributes `X`, `Y`

- `X`: protected attributes
- `Y`: other attributes
- `C`: binary outcome `{0,1}`

Direct discrimination is defined $C = f(X)$

Indirect discrimination is defined $C = f(Y)$

It's clear that X and Y are related. How do we proceed? Use the 80% rule:

If $P(C = 1 \mid X = 0) \leq 0.8\ P(C = 1 \mid X = 1)$ then we conclude disparate impact.

III. Checking for Disparate Impact

A. Idea: if we can predict the protected attribute X from the other attributes Y, this could lead to disparate impact

B. Metric: Balanced Error Rate (BER)

1. $BER = 1 / 2\ (P\ (f(Y)) = 0 \mid X = 1) + P\ (f(Y) = 1 \mid X = 0))$