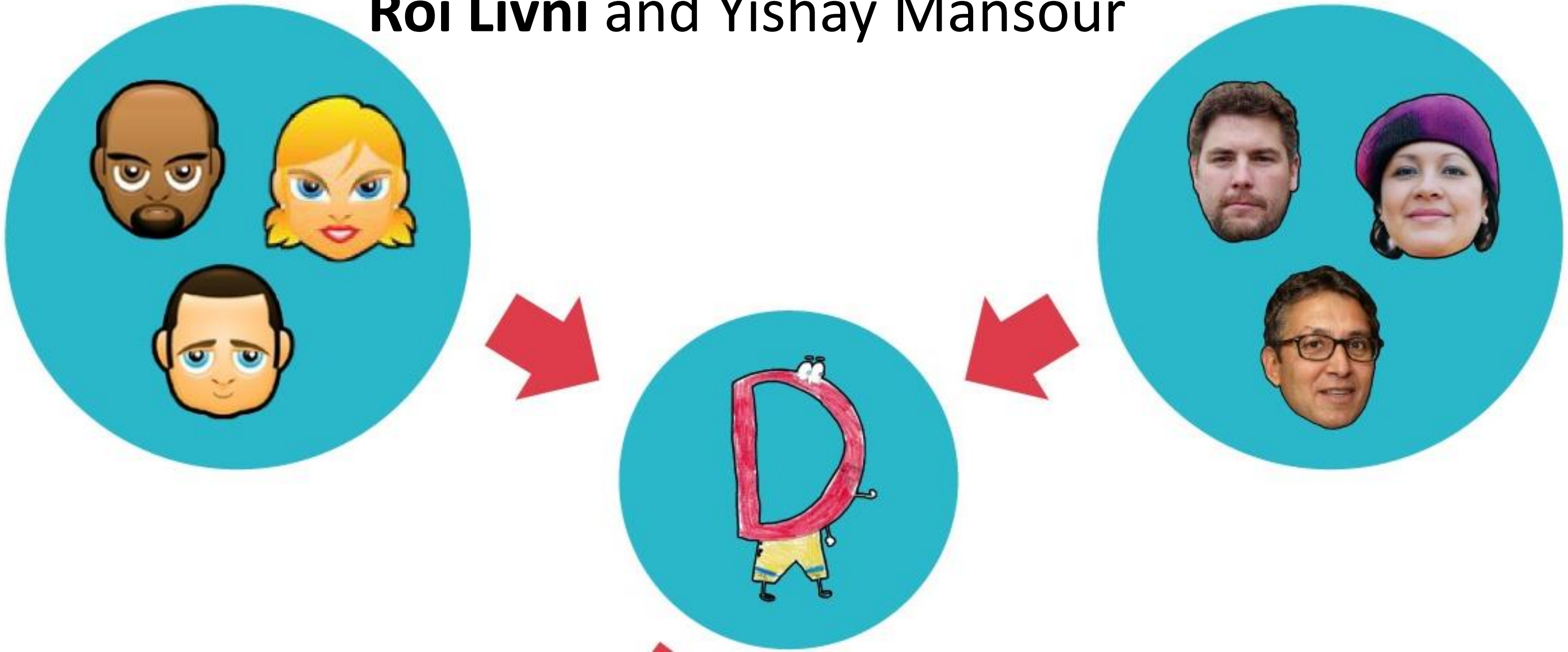
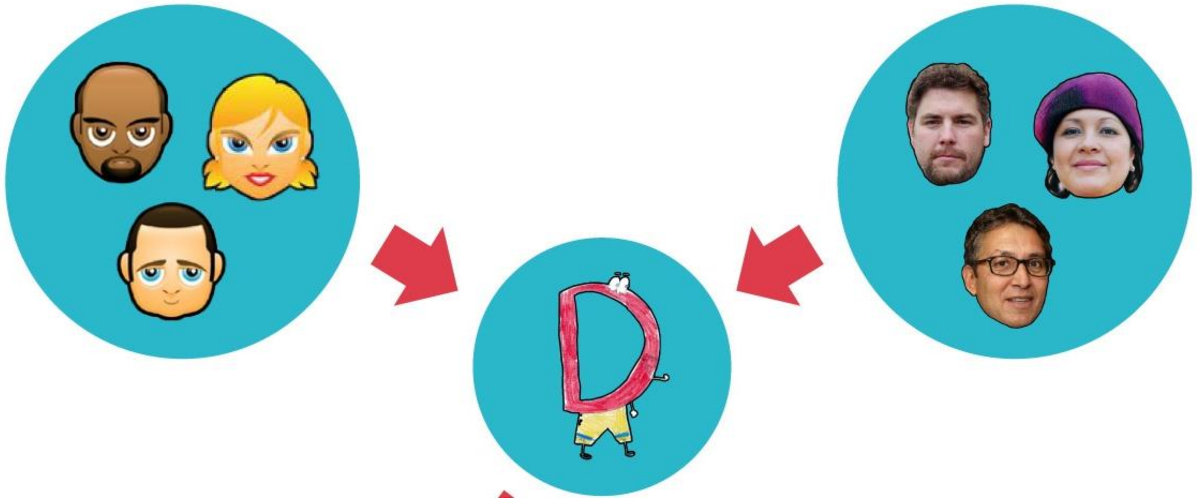
Graph Based- Discriminators Sample Complexity and Expressiveness

Roi Livni and Yishay Mansour

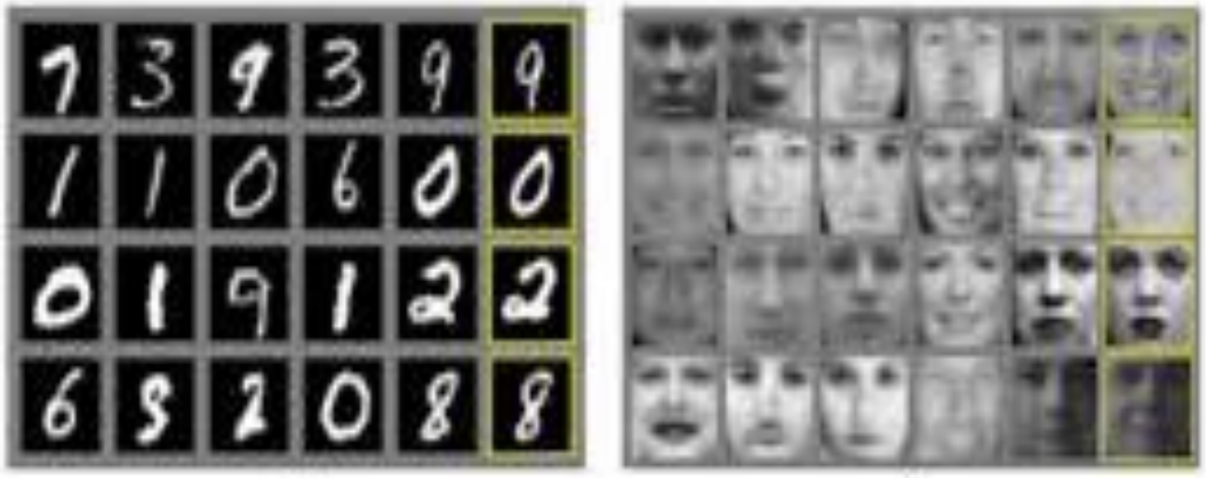


Discrimination

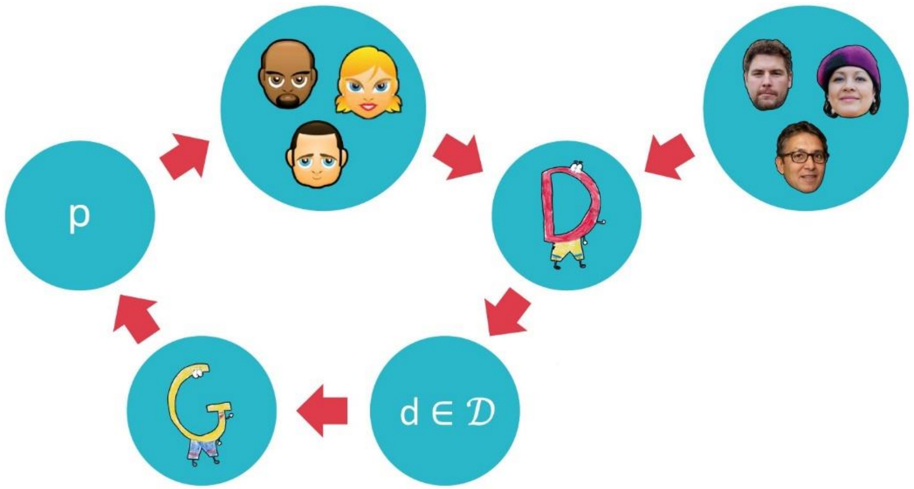
- A discriminator is provided with two data sets.
  - $S_1 \sim P_1$
  - $S_2 \sim P_2$
- Decide if P_1 and P_2 are different.
- If not, provide a certificate.



Motivation: Synthetic Data Generation



Goodfellow et al.'14



https://thispersondoesnotexist.com/

Discrimination: Learning Lens

Assume data is balanced: $P_1 = P(\cdot\,|y=1)$ $P_2 = P(\cdot|y=0)$.

- A learner is defined by a class $H \subseteq \{0,1\}^X$
- Given labelled sample from some distribution P over $X \times \{0,1\}$
- Learner returns $h \in H$ such that

$$P_{(x,y)}\left[h(x) \neq y\right] \leq \min_{h \in H} P_{(x,y)}\left[h(x) \neq y\right] + \epsilon$$

  - If $\sup_{h \in H} \left[E_{x\sim P_1}[h(x)] - E_{x\sim P_2}[h(x)]\right] > \epsilon$
    - Learner succeeds.

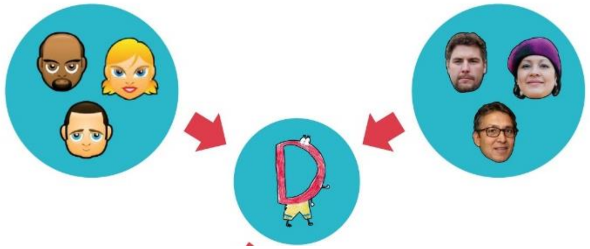
Learning as a discrimination task

- Discriminator is defined by a class of distinguishers $H \subseteq \{0,1\}^X$

Integral Probability Metric:
(Muller'97)

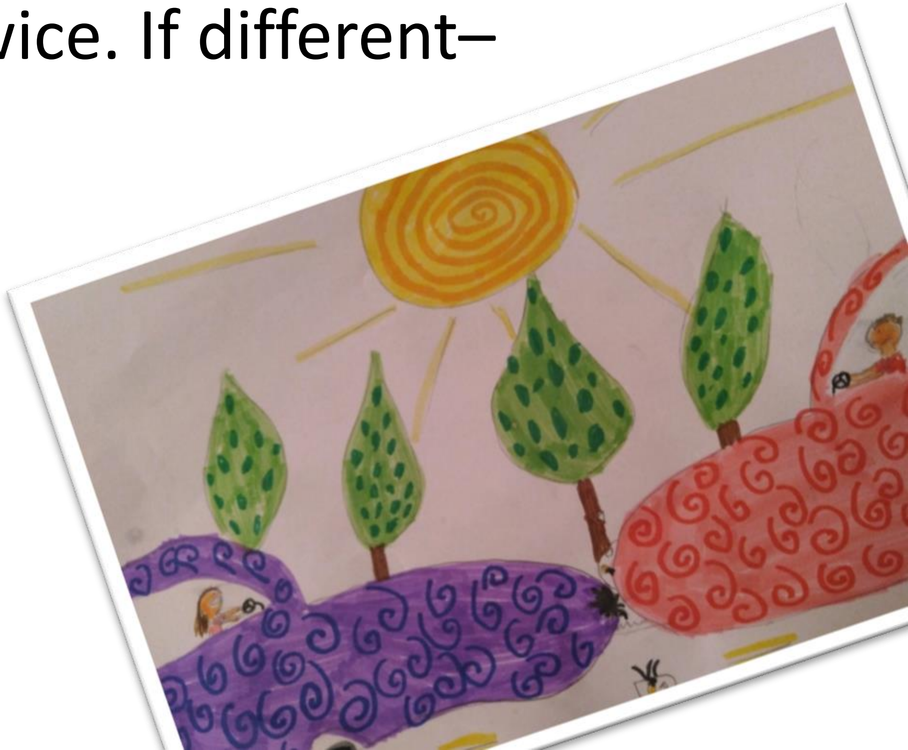
$$IPM_H(P_1, P_2) = \sup_{h \in H} |E_{x \sim P_1}[h(x)] - E_{x \sim P_2}[h(x)]|$$

- If $IPM_H(P_1, P_2) > \epsilon$ -- return $h \in H$ with $IPM_H(P_1, P_2) > \epsilon/2$
- If not, may fail. (return EQUIVALENT).



Higher order discrimination

- Instead of considering hypotheses classes, what if we take other types of statistical tests:
- **Example:** Collision test
- Estimate probability to draw the same point twice. If different– declare distinct.
- If not, may fail (return equivalent).



Higher order discrimination

- Instead of considering hypotheses classes, what if we take other types of distinguishers:
- **More generally:** Take a family $\mathrm{G} = \{g\colon\ g\colon X^2 \to \{0,1\}\}$

$$IPM_G(P_1, P_2) = \sup_{g \in G} \left| E_{(x_{1,}x_2)\sim P_1^2)}[g(x_1, x_2)] - E_{(x_{1,}x_2)\sim P_2^2)}[g(x_1, x_2)] \right|$$

- Are graph-based distinguishers stronger than classical distinguishers?
- Sample Complexity

Expressive power of graph-based discriminators

THEOREM: Let X be an infinite domain. There exists a graph g such that: For every hypothesis class H with finite VC dimension and $\epsilon > 0$, there are two distributions P_{syn}, P_{real} such that

$$IPM_H(p_{syn}, p_{real}) < \epsilon$$

$$and,$$

$$\left| E_{(x_1,x_2)\sim p^2_{syn}[g(x_1,x_2))]} - E_{(x_1,x_2)\sim p^2_{real}}[g(x_1,x_2)] \right| > \frac{1}{4}$$

(L, Mansour'19)

Finite Version

- If |X|=N, there is a graph g such that for every class H there are two distributions that are H-indistinguishable, g-distinguishable unless:
  - $VC(H) = \Omega(\epsilon^2 \log N)$ (L, Mansour'19)

- **Optimal**: For every graph-based class G with finite capacity there is a hypothesis class H with VC dimension $O(\epsilon^2 \log N)$ such that

$$IPM_C(p_{syn}, \mathrm{p}_{real}) > \frac{1}{4} \Rightarrow IPM_G(p_{syn}, \mathrm{p}_{real}) > \epsilon$$

(Alon, L, Mansour)

Sample complexity of graph-based discriminators

- For a family of graph G.
- Given samples from two unknown distributions P_1, P_2: Decide if

$$IPM_G(P_1, P_2) > \epsilon$$

- How many examples are needed?

- Recall:

  - For an hypothesis class, a discriminator can decide if $IPM_H(P_1, P_2) > \epsilon$, if and only if H has finite VC dimension.

  - $\Theta(VC(H)/\epsilon^2)$ are needed

The graph-VC dimension

- The graph VC dimension is obtained by considering the projections of the graph by fixing a vertex. Namely, for every x consider the hypothesis class

$$H_x = \{g(x,\cdot): X \to \{0,1\}: g \in G\}$$

- Then: $gVC\ (C) = \sup_{x \in X} VC(H_x)$

- $O(gVC(C))$ are sufficient.
- $\Omega(\sqrt{gVC(C)})$ are necessary.

(L, Mansour'19)