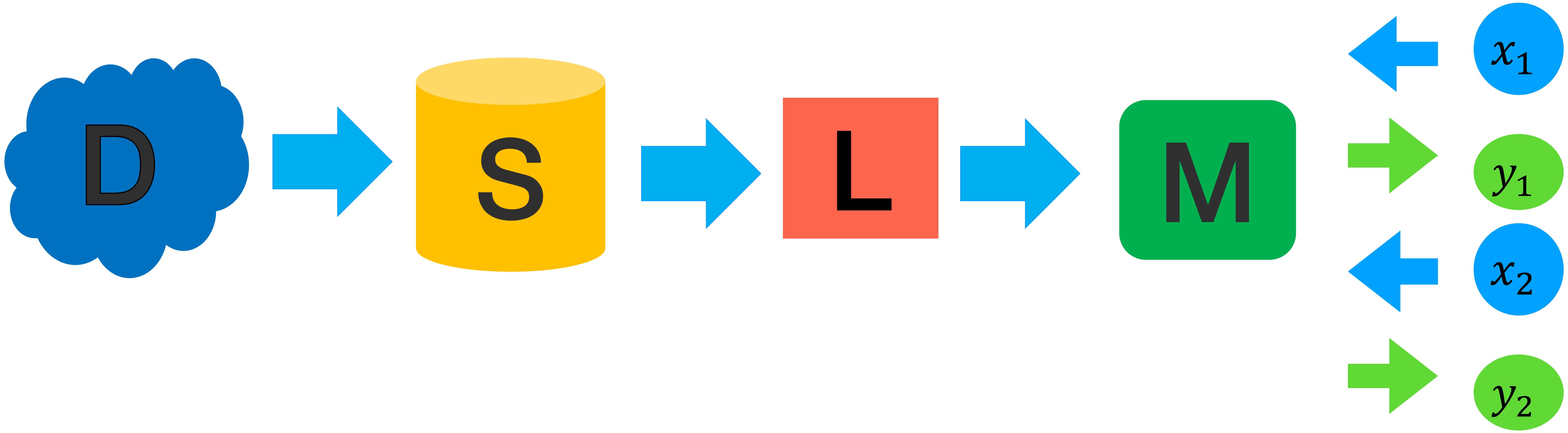
Private Everlasting Prediction

Moni Naor[1], Kobbi Nissim[2], Uri Stemmer[3], Chao Yan[2]

1. Weizmann Institute of Science
2. Georgetown University
3. Tel Aviv University

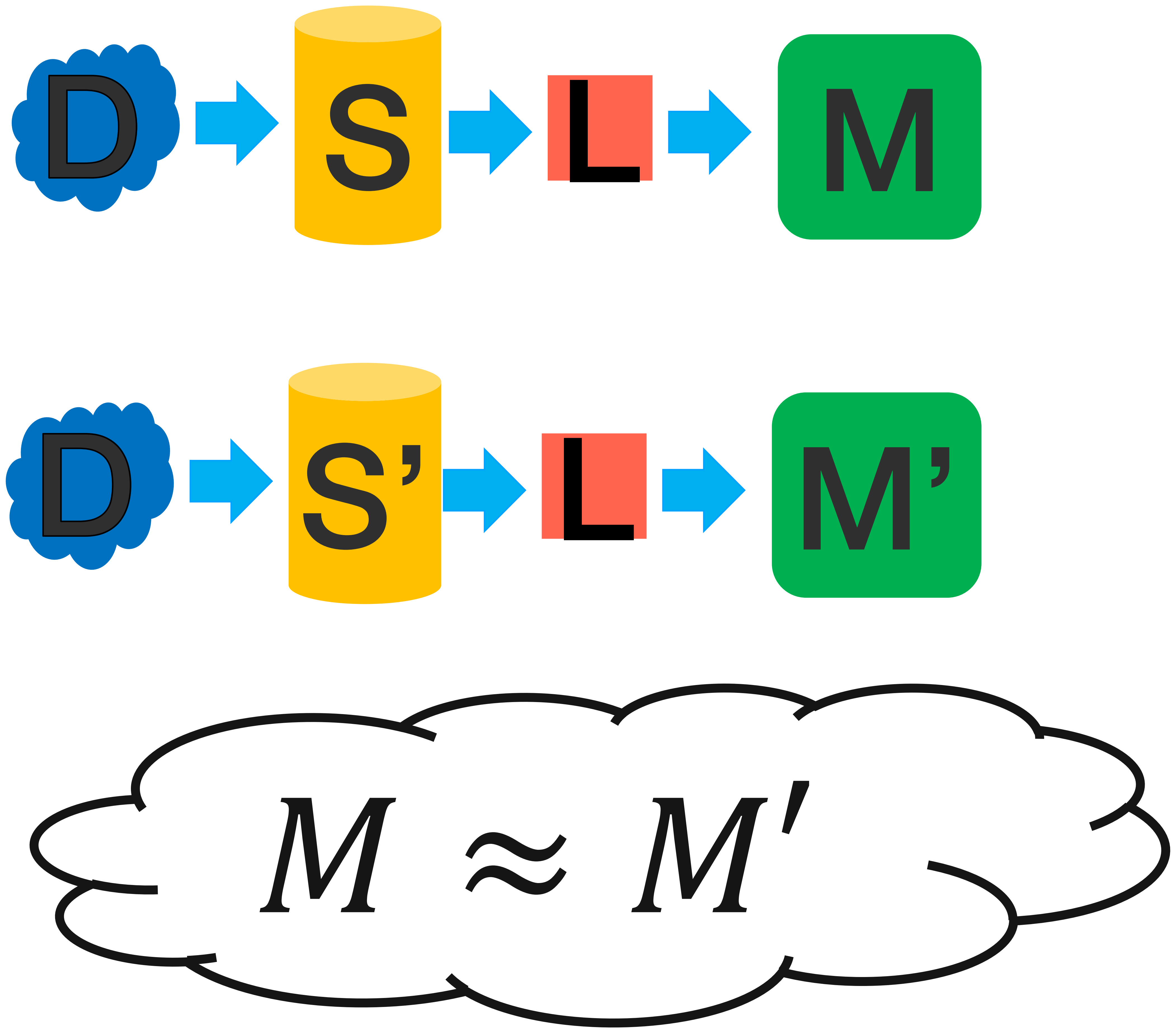
How many examples do we need?



PAC learning [Valiant 84] : $\Theta(VC(C))$

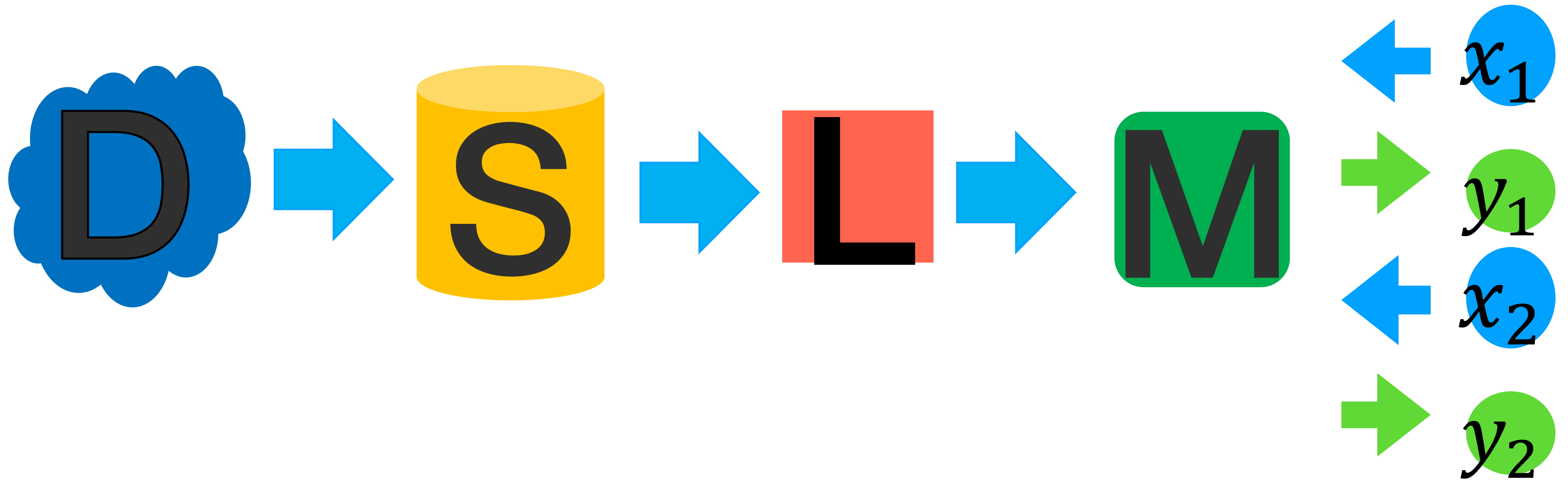
Private PAC learning [KLNRS 08]

- Learner PAC learns the model
- Learner L is differentially private
- Theorem: n examples suffice,
  - $n = O(\log|C|)$ [KLNRS 08]
  - $n = O((LDim|C|)^6)$ [BLM20, GGKM20]



Price of private learning

- PAC learning: $n = \Theta(VC(C))$
- Private learning: $n = O(\min(\log|C|, LDim^6(C))$



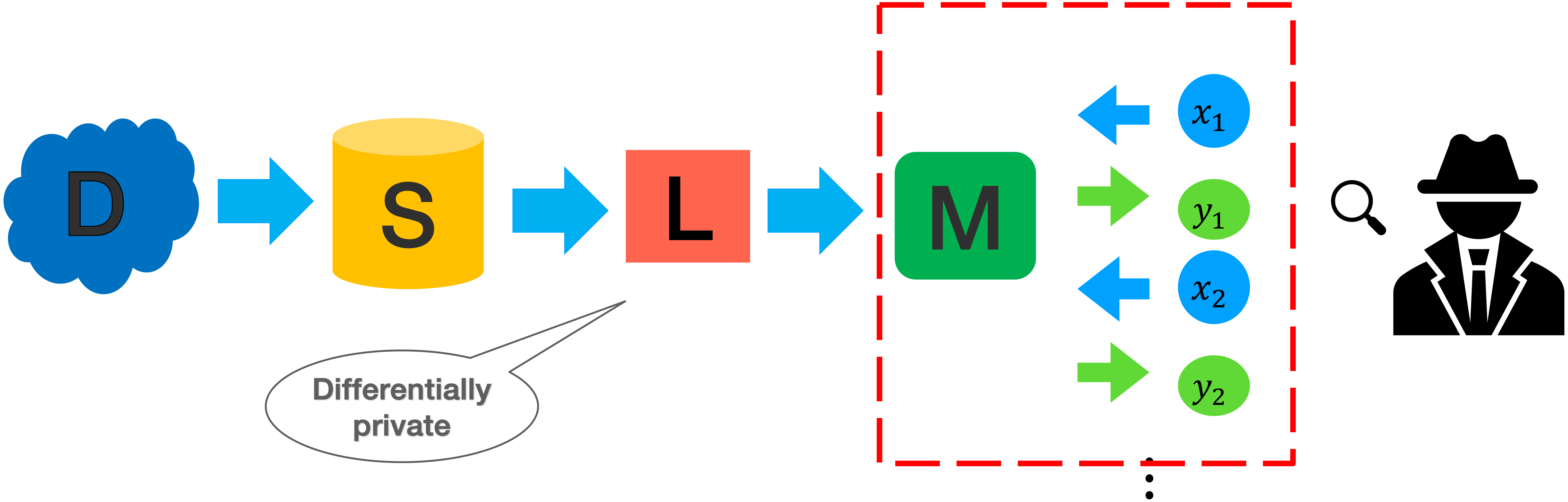
- $VC(C) \leq \min(\log|C|, LDim(C))$

Learning threshold functions		
PAC learning	Pure DP learning	Approximate DP learning
$n = O(1)$	$n = \Theta(\log|X|)$ [FX 15]	$n = \Theta(\log^*|X|)$ [BNS 13, BNSV 15, CLNSS 23,…]

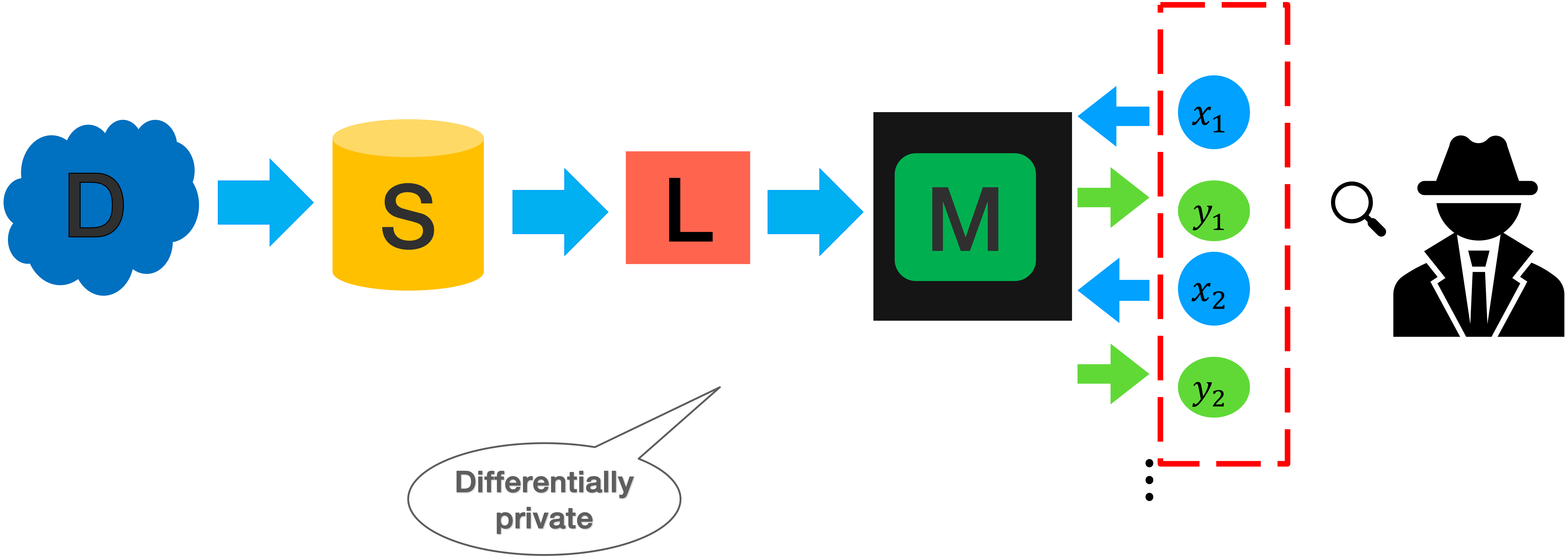
n grows with $|X|$

☹

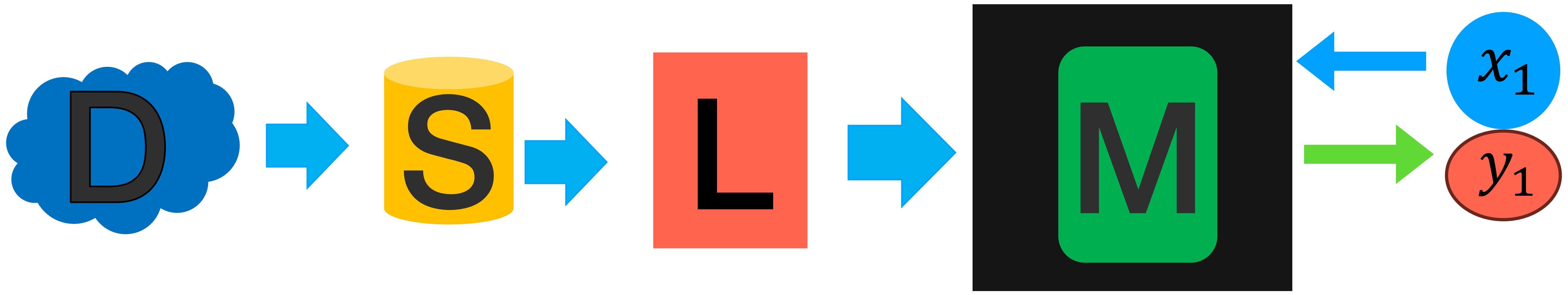
Rethinking private learning



Rethinking private learning



Blackbox prediction [Dwork Feldman 18]



- Labeled dataset $S = ((x_1, y_1), \dots, (x_n, y_n))$
- A single differentially private prediction: on query x, output the label y
- $n = \Theta(VC(C))$
- Answer t prediction queries by increasing n to $O\left(\sqrt{t}VC(C)\right)$ (using advanced composition)

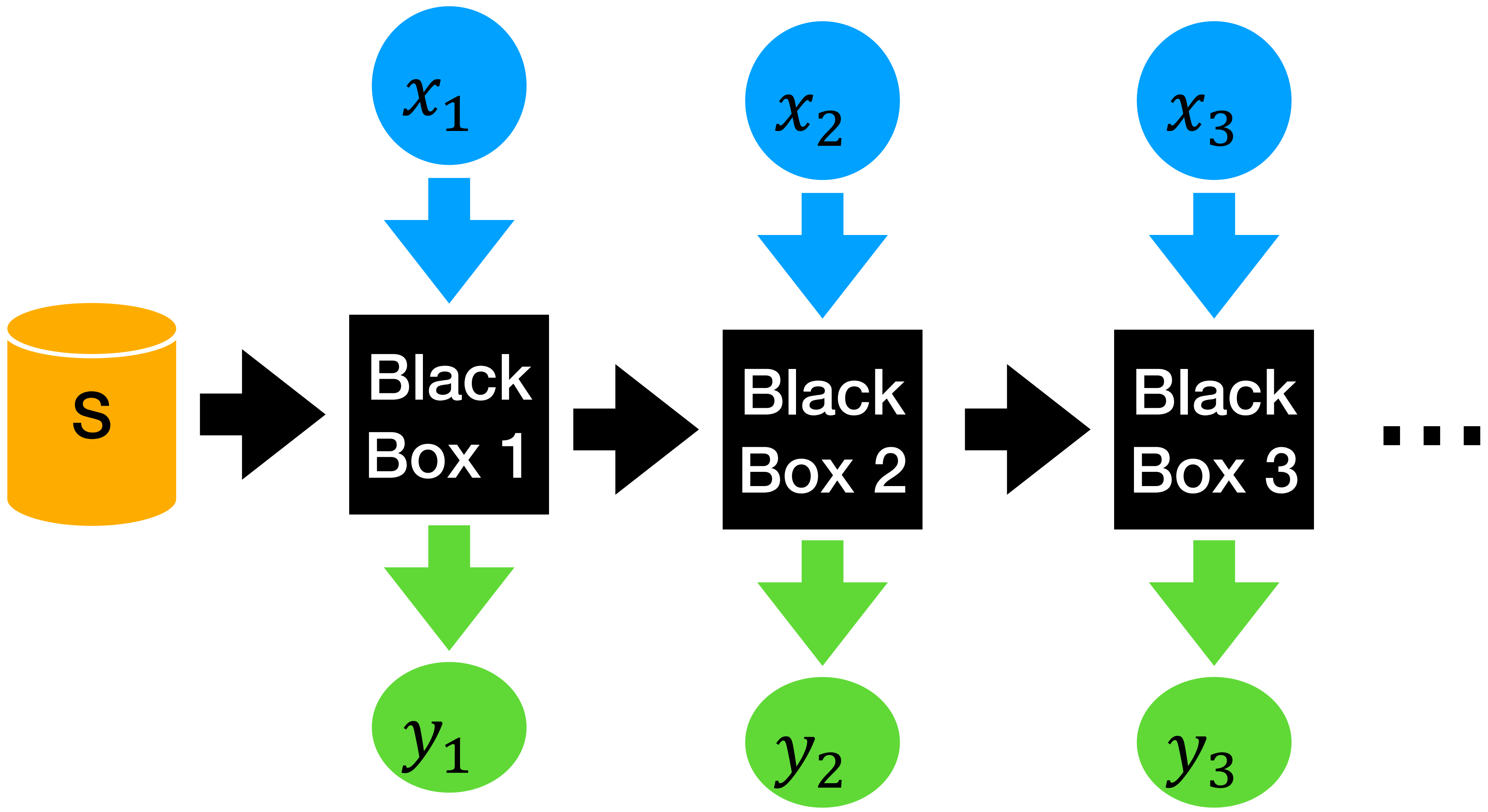
Main Result: private everlasting prediction

Predict an unlimited number of queries

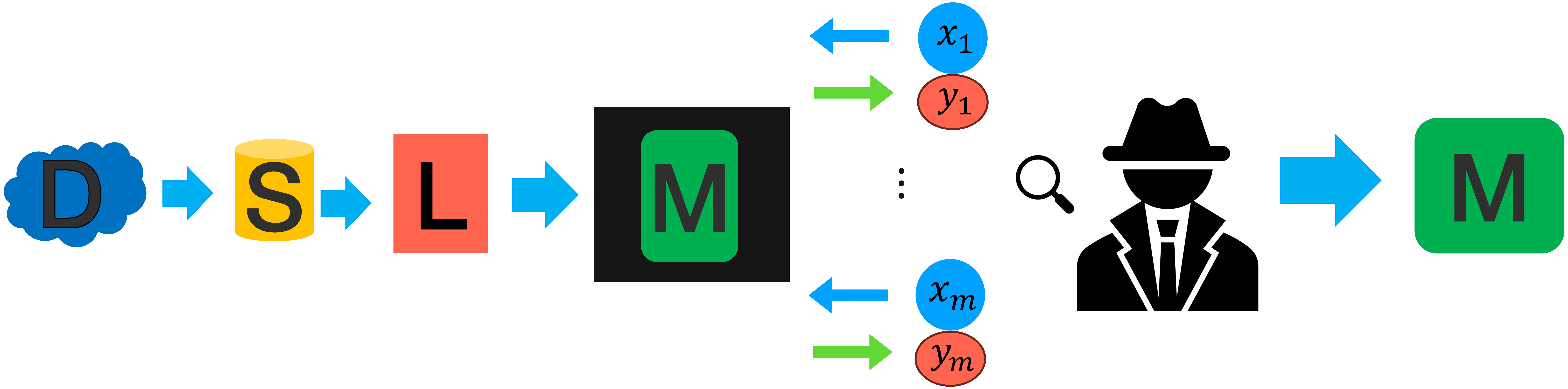
- Given labeled dataset $S = ((x_1, y_1), \ldots, (x_n y_n))$, we can privately predict an unlimited number of queries, where $n = O(VC^2(C))$.

- Utility Guarantee: with probability 1-β, every query is answered with α-accurate hypothesis

- Privacy guarantee: differentially private both for S and queries. An adaptive version of JDP [Kearns et al. 2015]

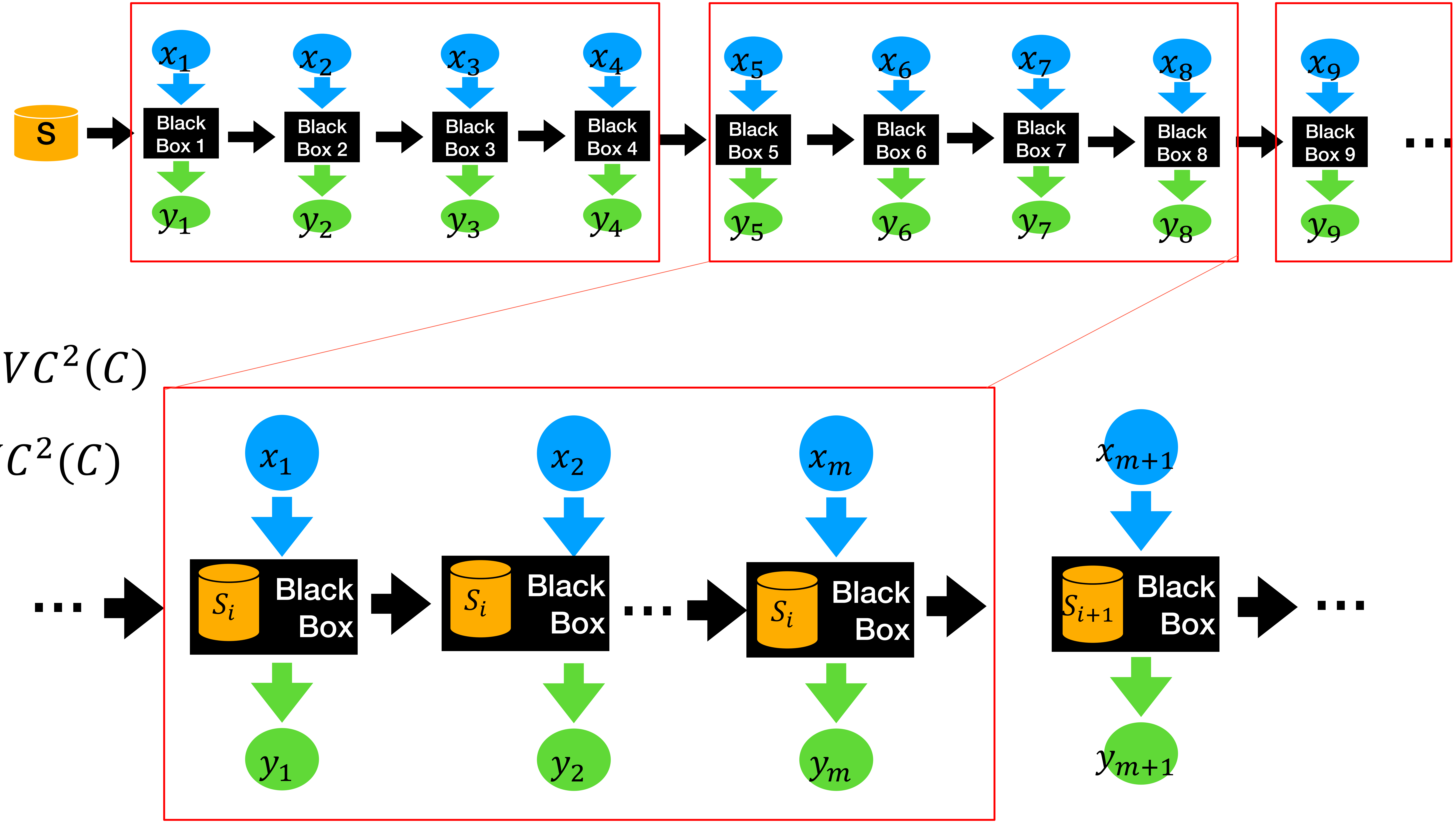


Observation: black box prediction cannot only depend on S



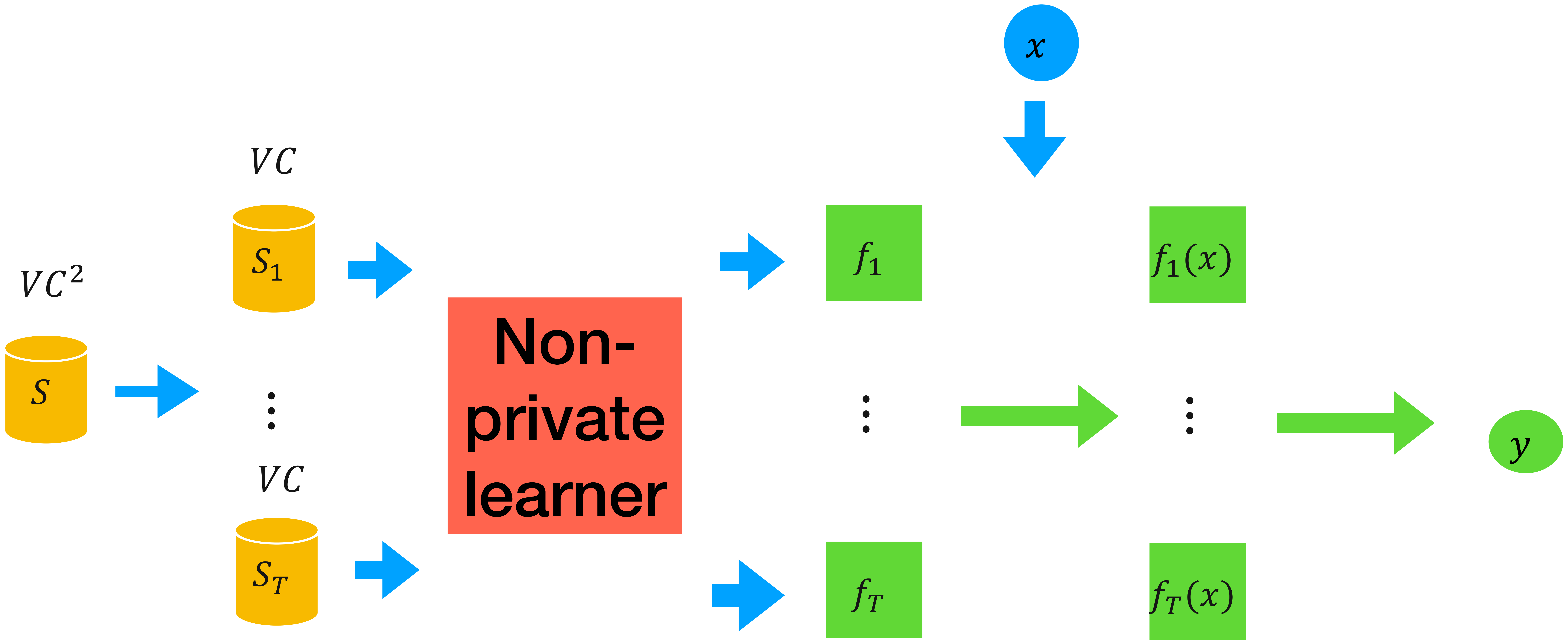
- One black box private prediction implies private learning

A generic construction



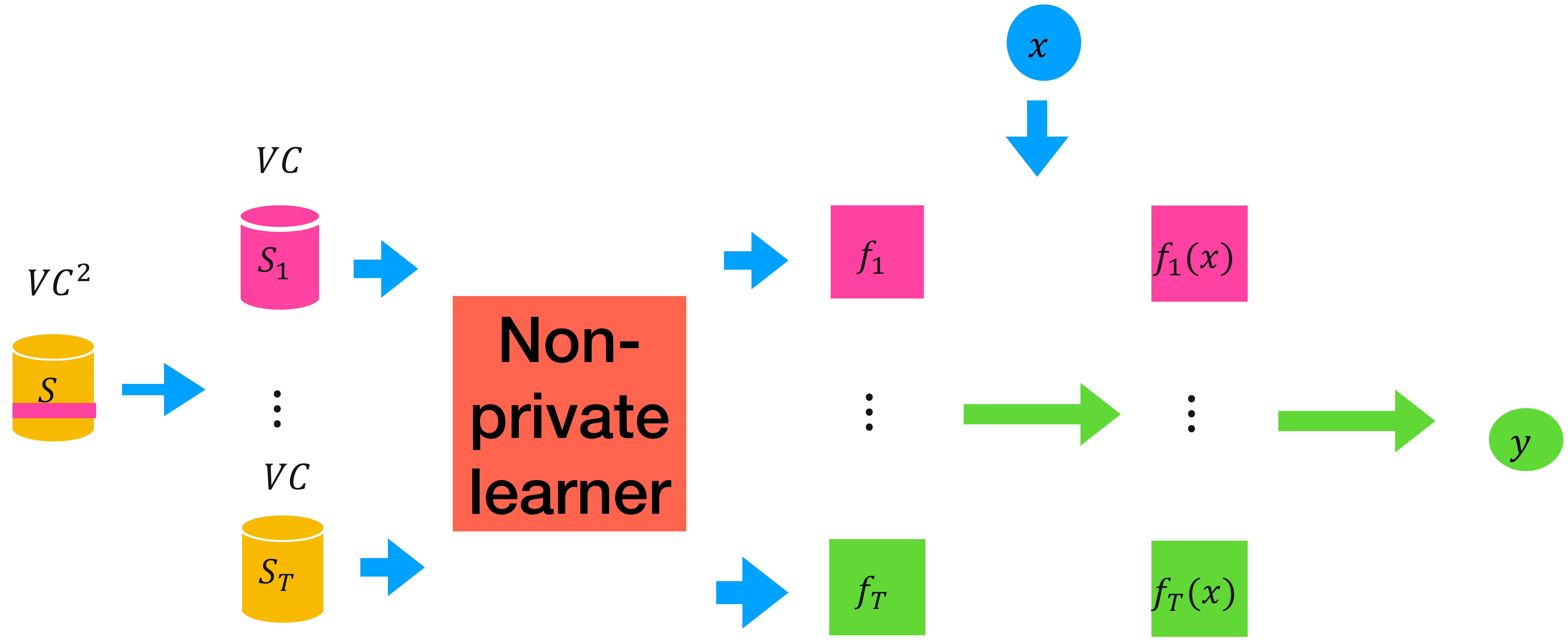
- $|S_i| \approx VC^2(C)$
- $m \approx VC^2(C)$

Prediction queries



$$y = noisy\ maj(f_1(x), \dots, f_T(x))$$

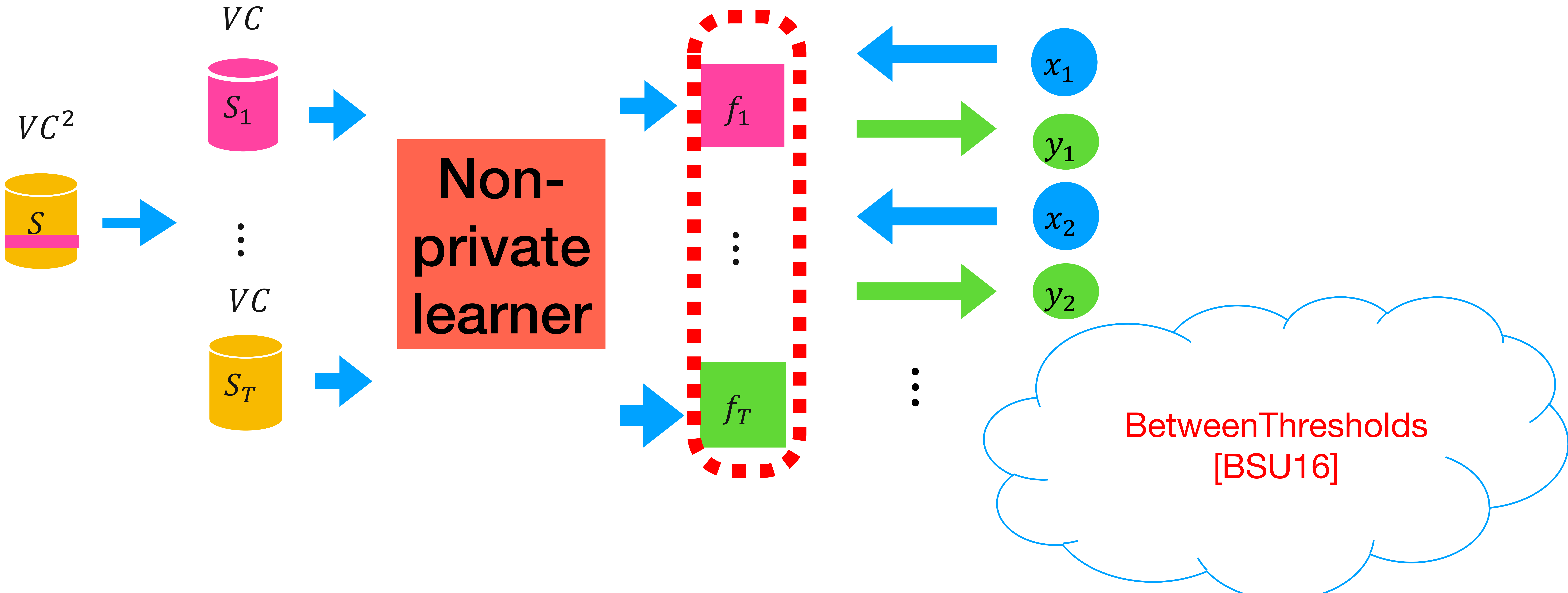
Privacy of labeled set S



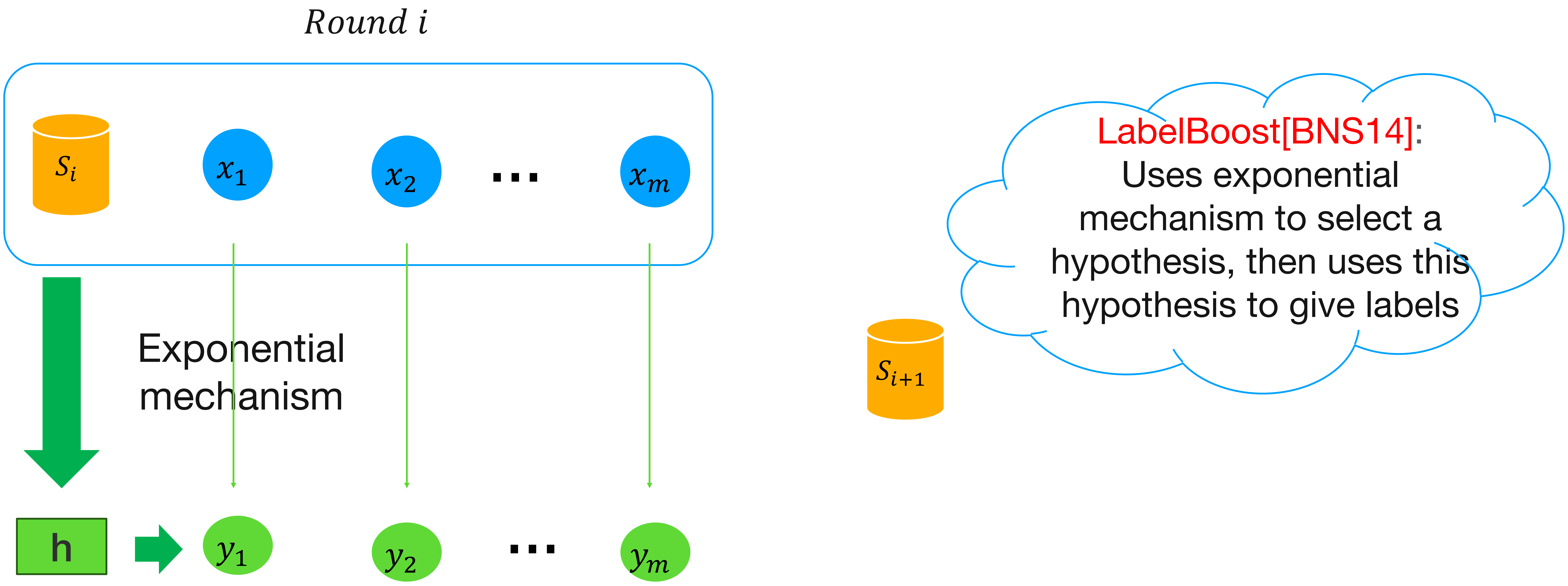
$$y = noisy\ maj(f_1(x), \dots, f_T(x))$$

Privacy of labeled set S

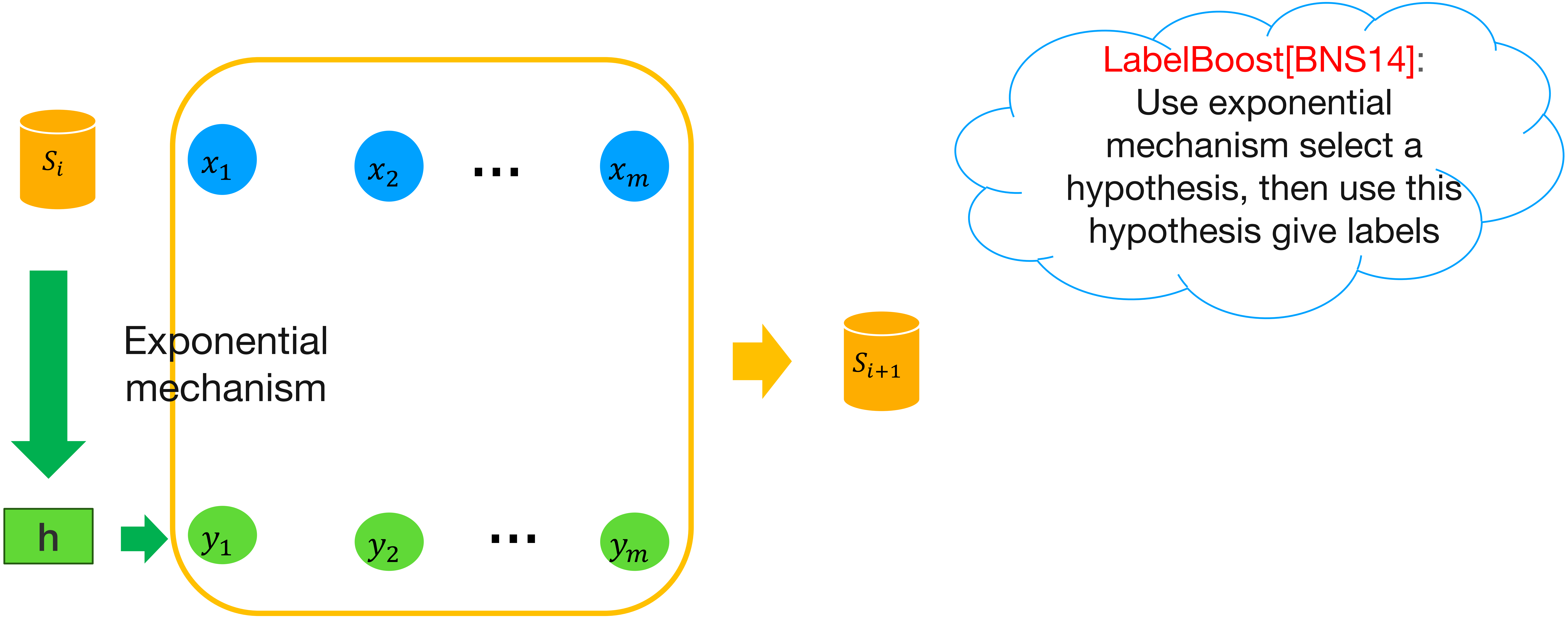
$$y = noisy\ maj(f_1(x), \dots, f_T(x))$$



Generating labeled samples for the next round



Update S_i



S_i
x_1
x_2
…
x_m
Exponential mechanism
h
y_1
y_2
…
y_m
S_{i+1}
LabelBoost[BNS14]:
Use exponential mechanism select a hypothesis, then use this hypothesis give labels

Summary

- Everlasting prediction alternative to private learning
  - Predict any concept class with finite VC (e.g. thresholds over the reals)
  - It is efficient on some hard tasks for private learning (e.g. EncThresh [BZ15])

	Private learning	Our work
Thresholds over reals	impossible	$n = O(1)$
EncThresh	Time inefficient	Time efficient

- Open questions
  - Could $|S|$ be reduced to linear in VC?
    - It's VC^2 in our construction
  - Could this construction be made polynomial time?
    - We use exponential mechanism to generate new dataset.

Thank you