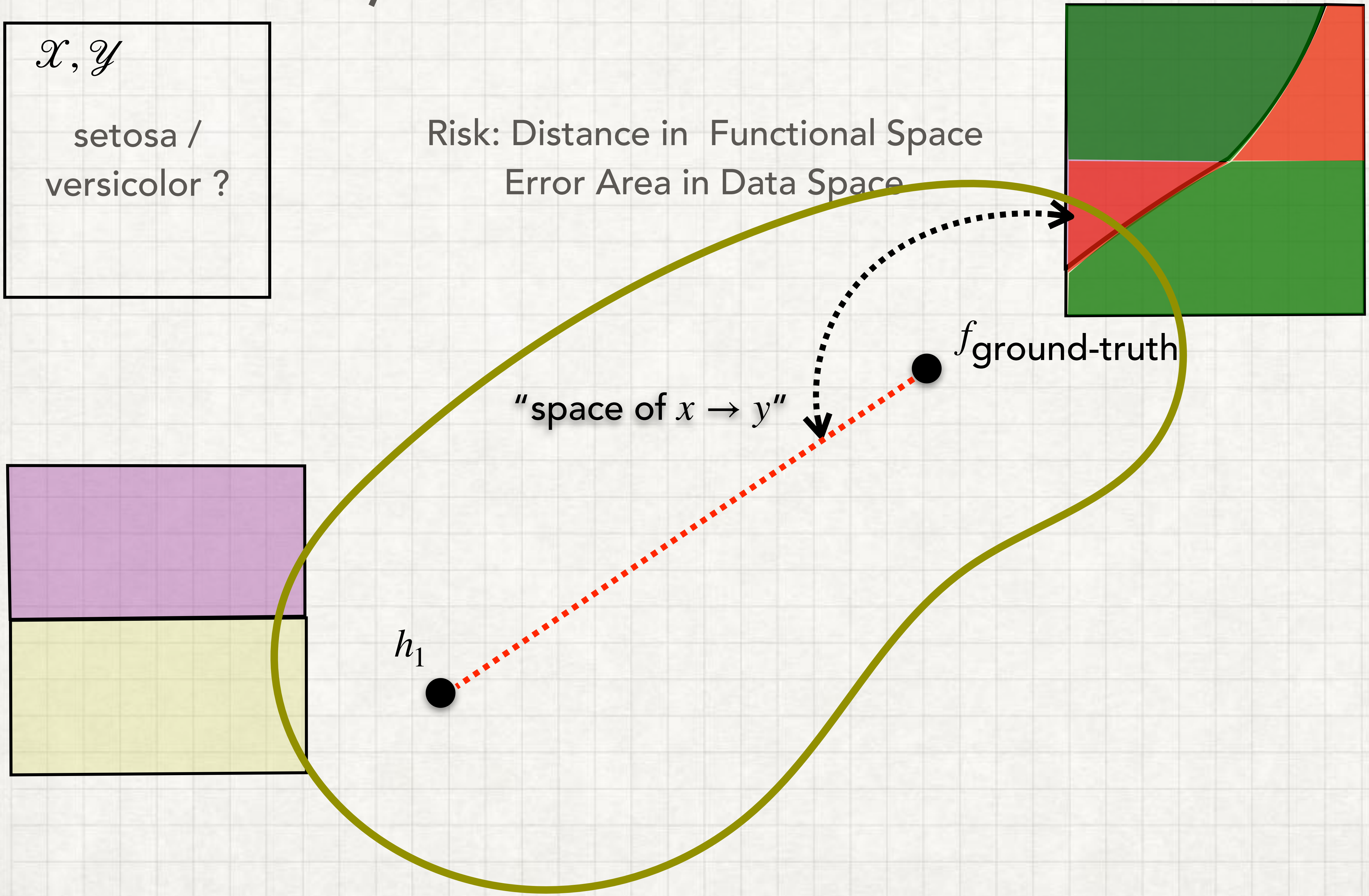
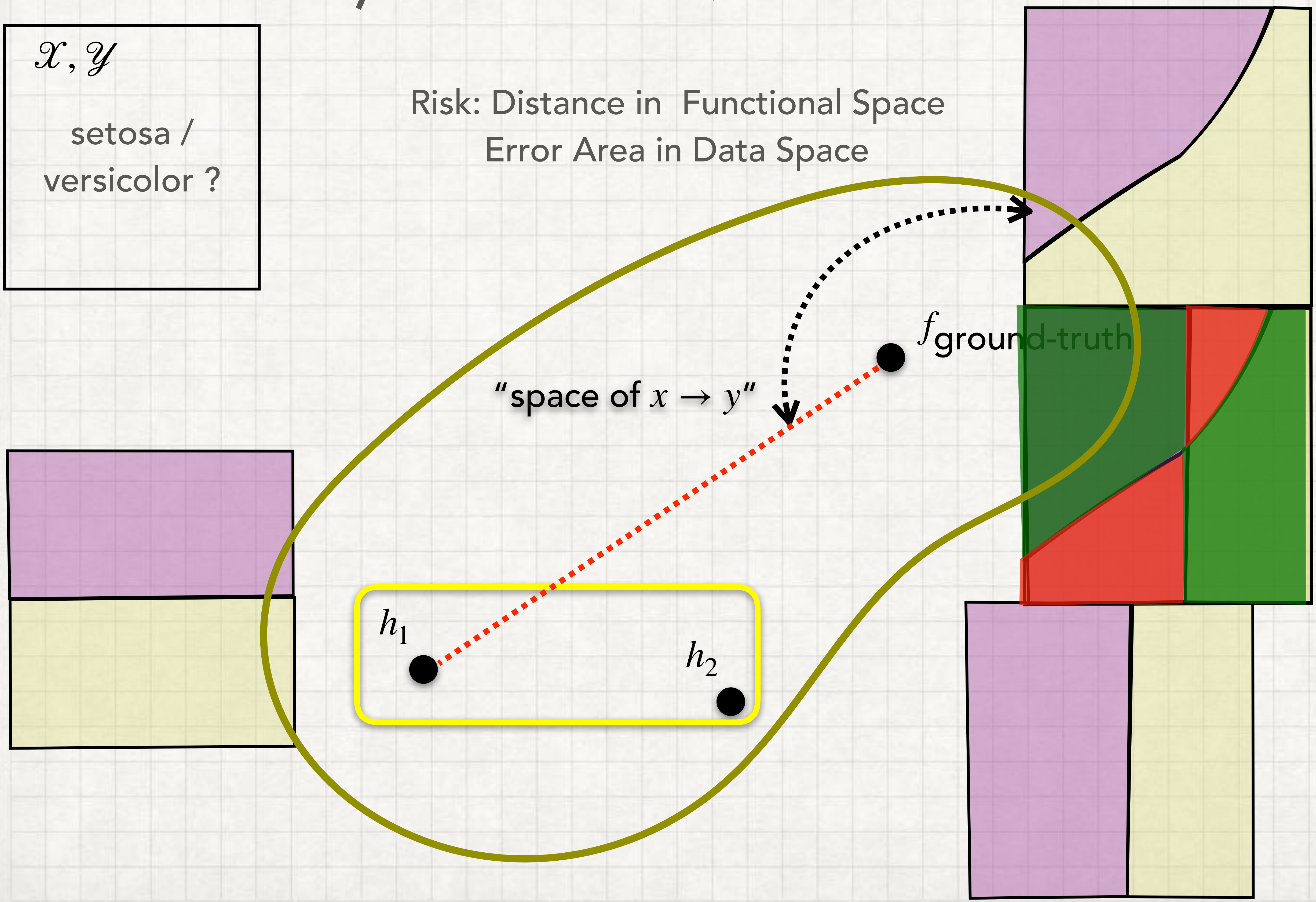
BIAS VARIANCE BOOTSTRAPPING AND BOOSTING

MOD1: INTUITIVE INTERPRETATION OF $\mathscr{H}$ FAMILY AND RISK

ERROR/RISK IN FUNCTIONAL SPACE



ERROR/RISKS OF TWO HYPOTHESES

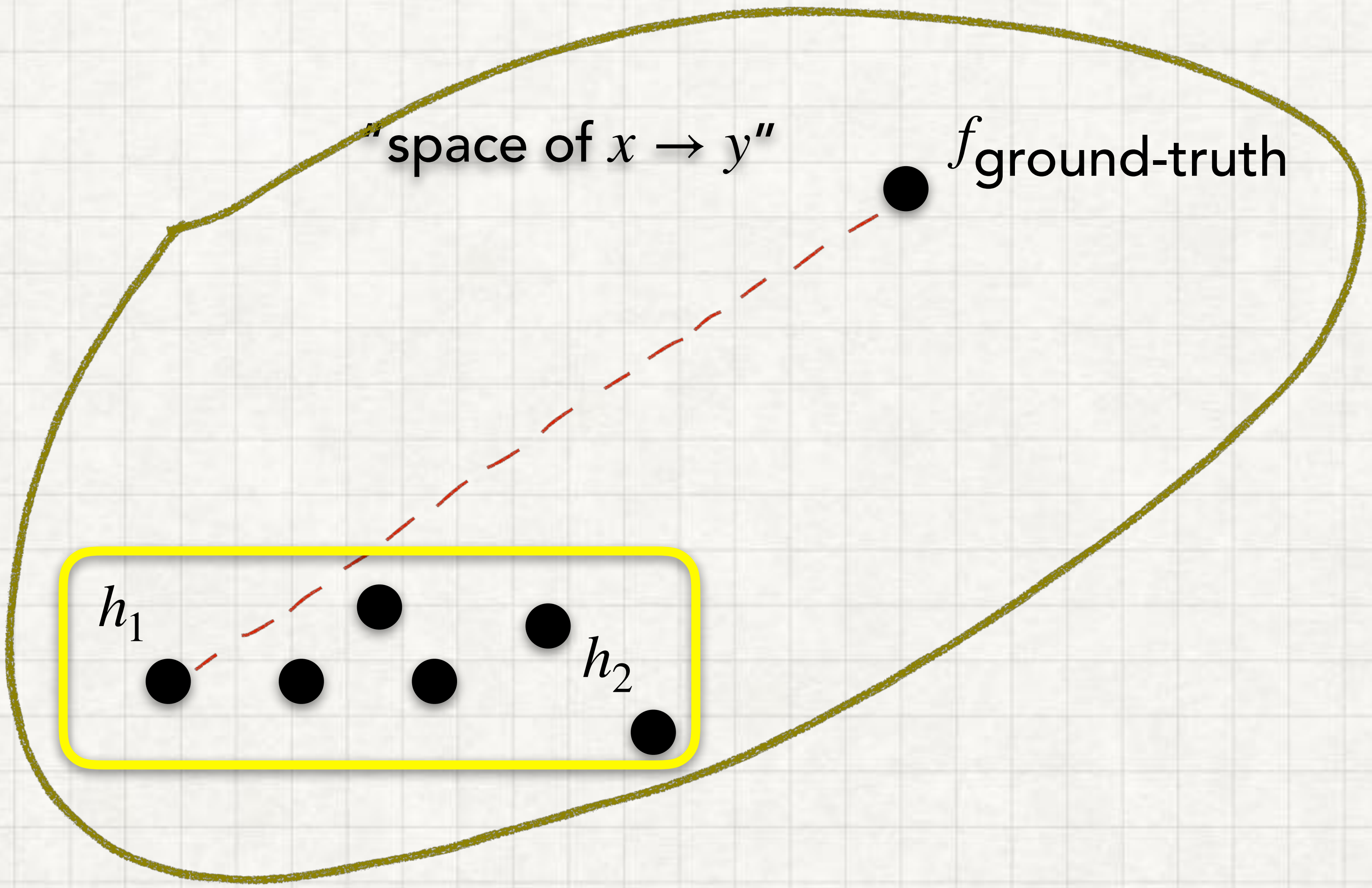


ERROR/RISKS OF MANY HYPOTHESES

$\mathcal{X}, \mathcal{Y}$

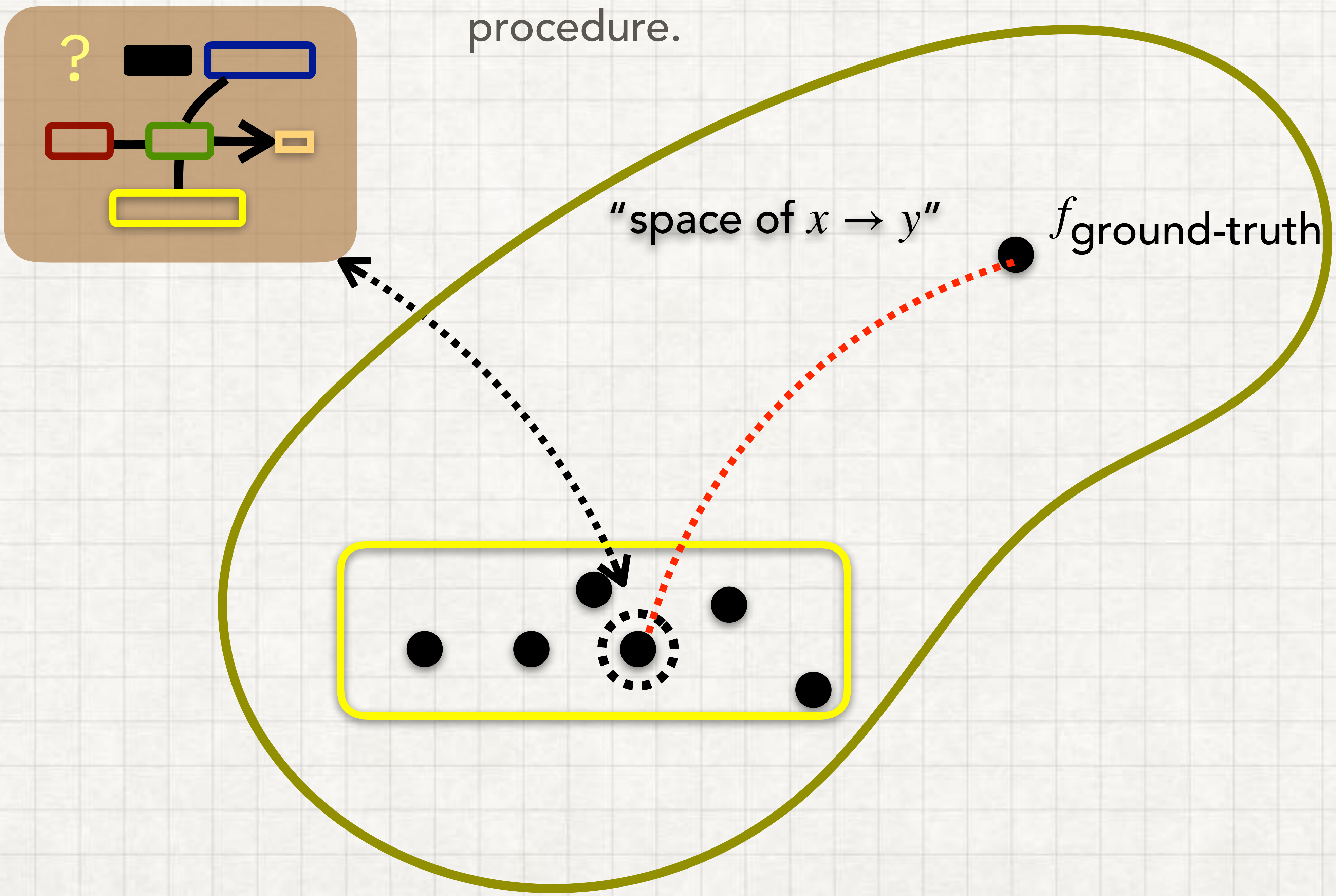
Risk: Distance in Functional Space

Error Area in Data Space

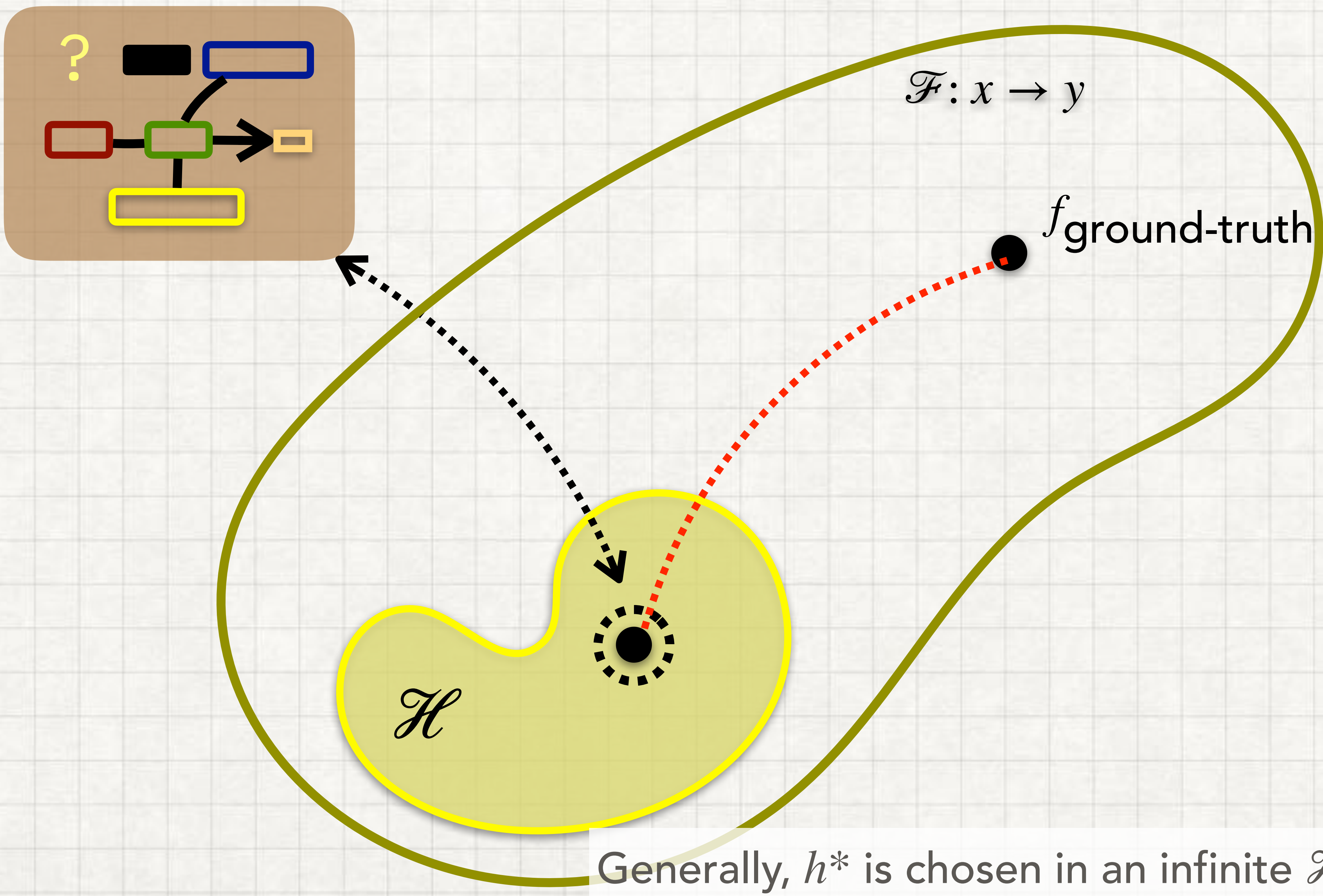


ERROR/RISKS OF SELECTED HYPOTHESES FROM MANY

The learning problem is concerned with the hypothesis selected by the learning procedure.



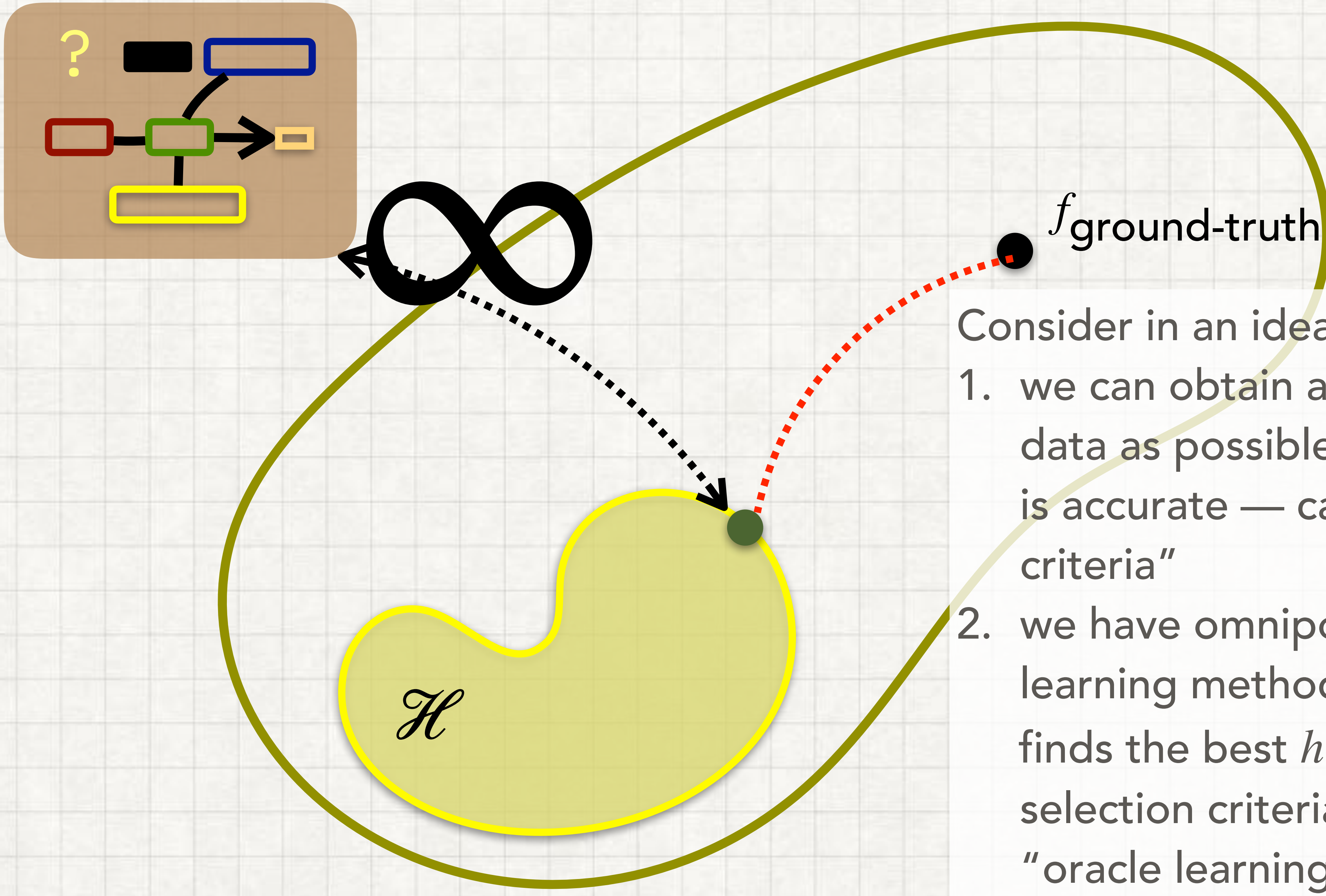
ERROR/RISKS OF SELECTED HYPOTHESIS FROM $\mathscr{H}$



Generally, h^* is chosen in an infinite $\mathscr{H}$.

MOD2: DISCRIMINATE TWO TYPES OF RISKS WHEN FACING RANDOM DATA BIAS AND VARIANCE

OPTIMUM WITHIN $\mathscr{H}$

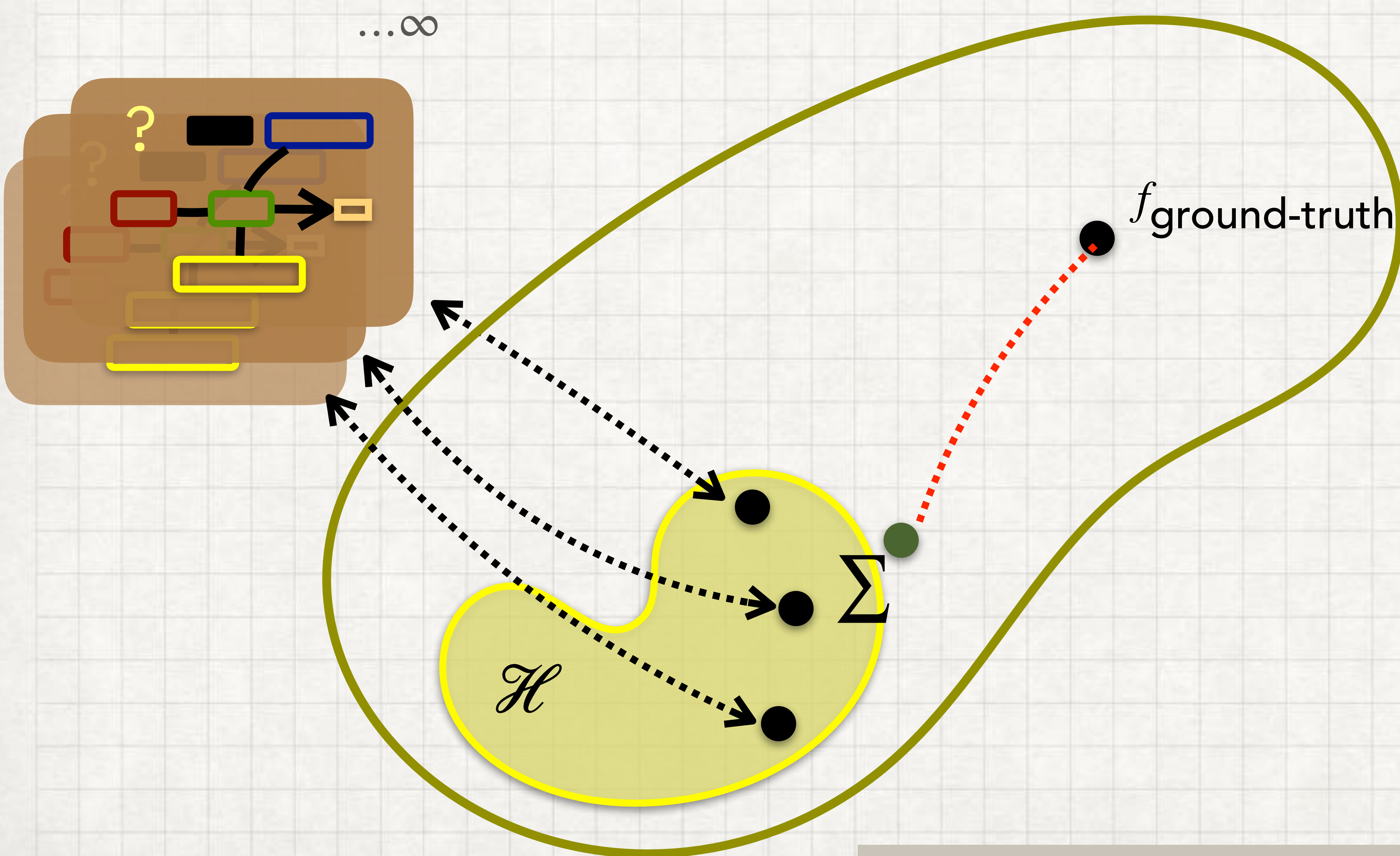


Consider in an ideal world,

1. we can obtain as many training data as possible — the evaluation is accurate — call this "oracle criteria"
2. we have omnipotently powerful learning method, which always finds the best $h^* \in \mathscr{H}$ for any selection criteria — call this "oracle learning algorithm"

Putting together, we then consider the minimal risk of an $\mathscr{H}$.

NOT INFINITE DATA, BUT INFINITE LEARNING EXPERIMENTS AND AVERAGING



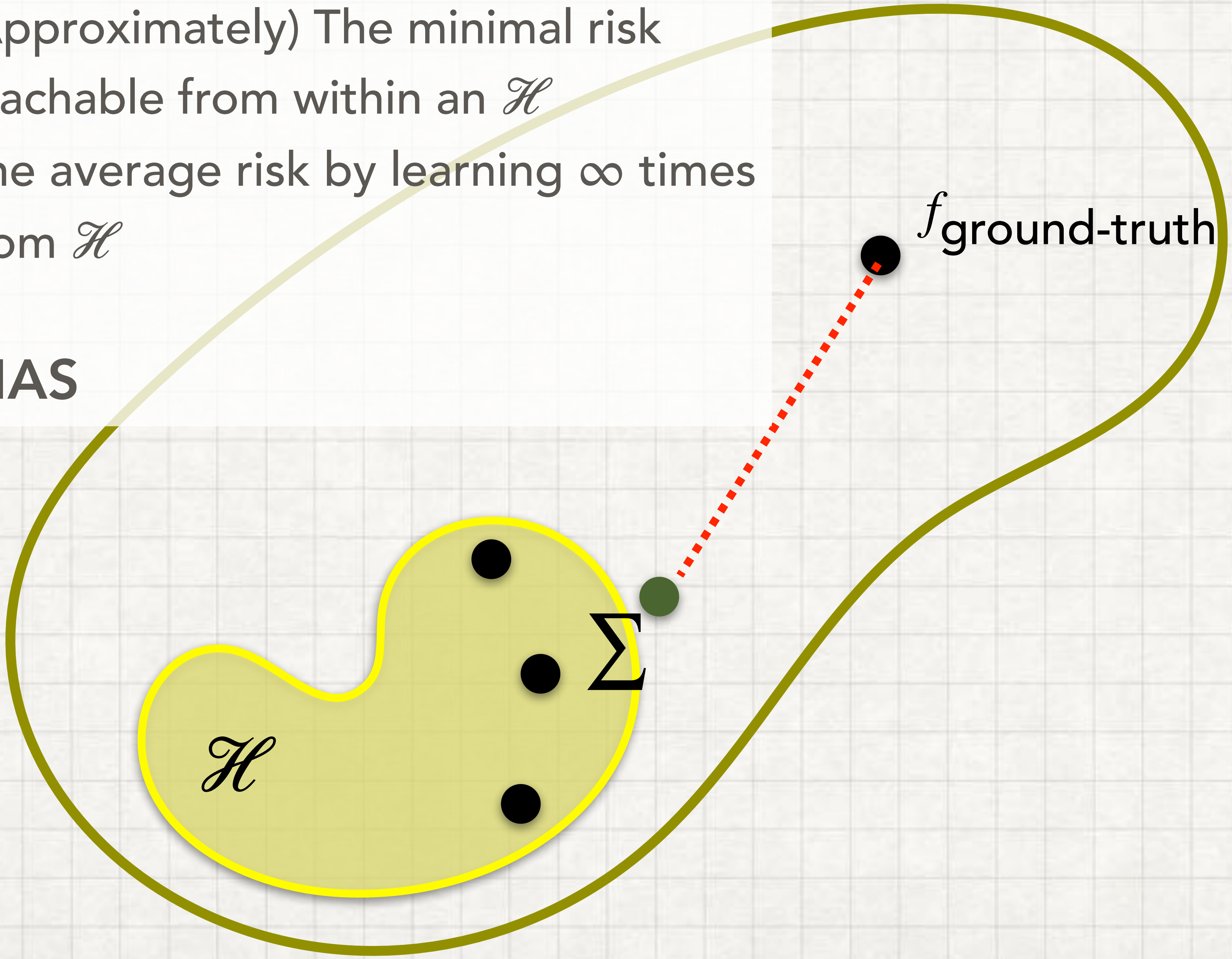
The average predictor is not always within the original $\mathscr{H}$

DEFINITION OF BIAS

(Approximately) The minimal risk reachable from within an $\mathscr{H}$

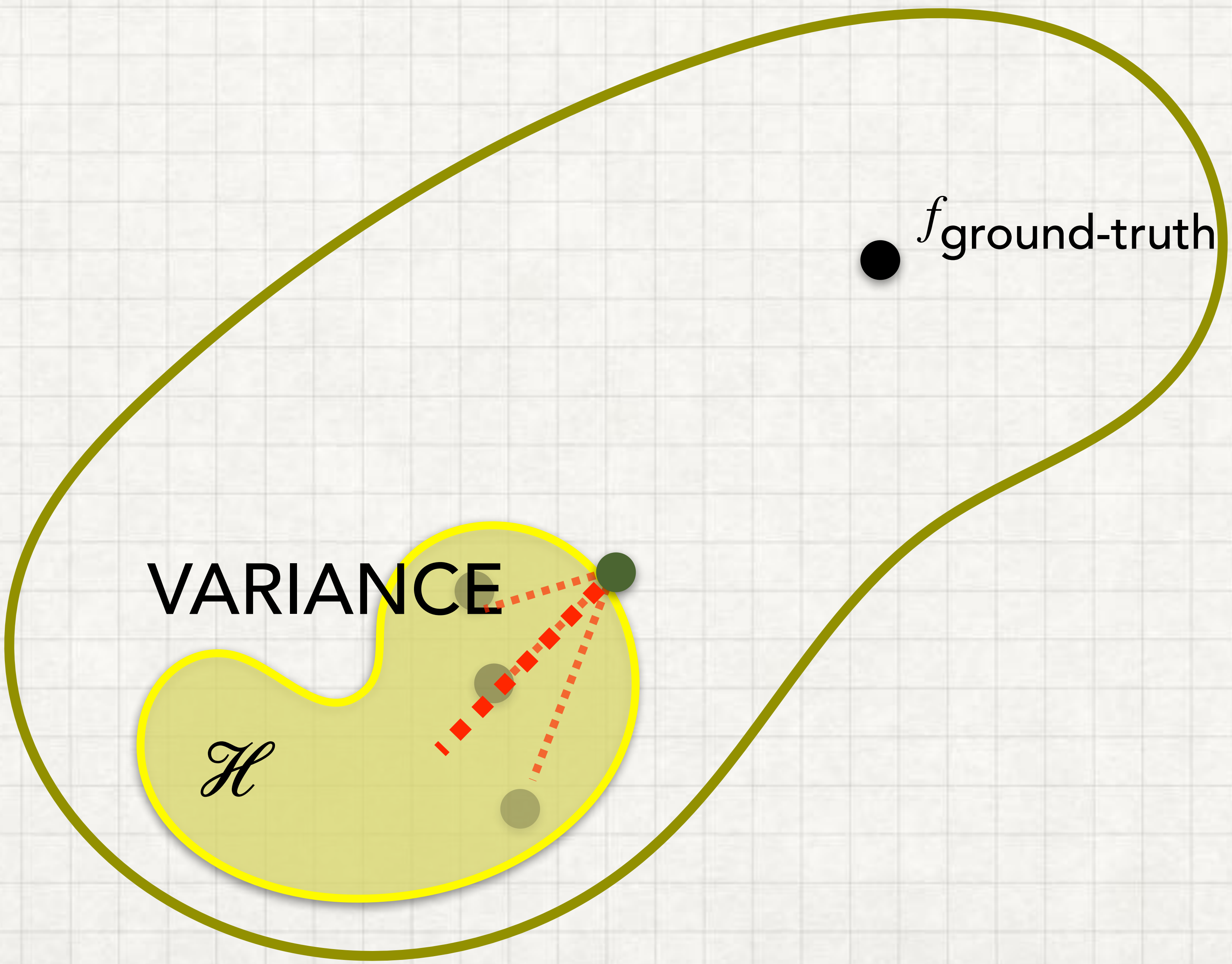
The average risk by learning ∞ times from $\mathscr{H}$

∞ **BIAS**



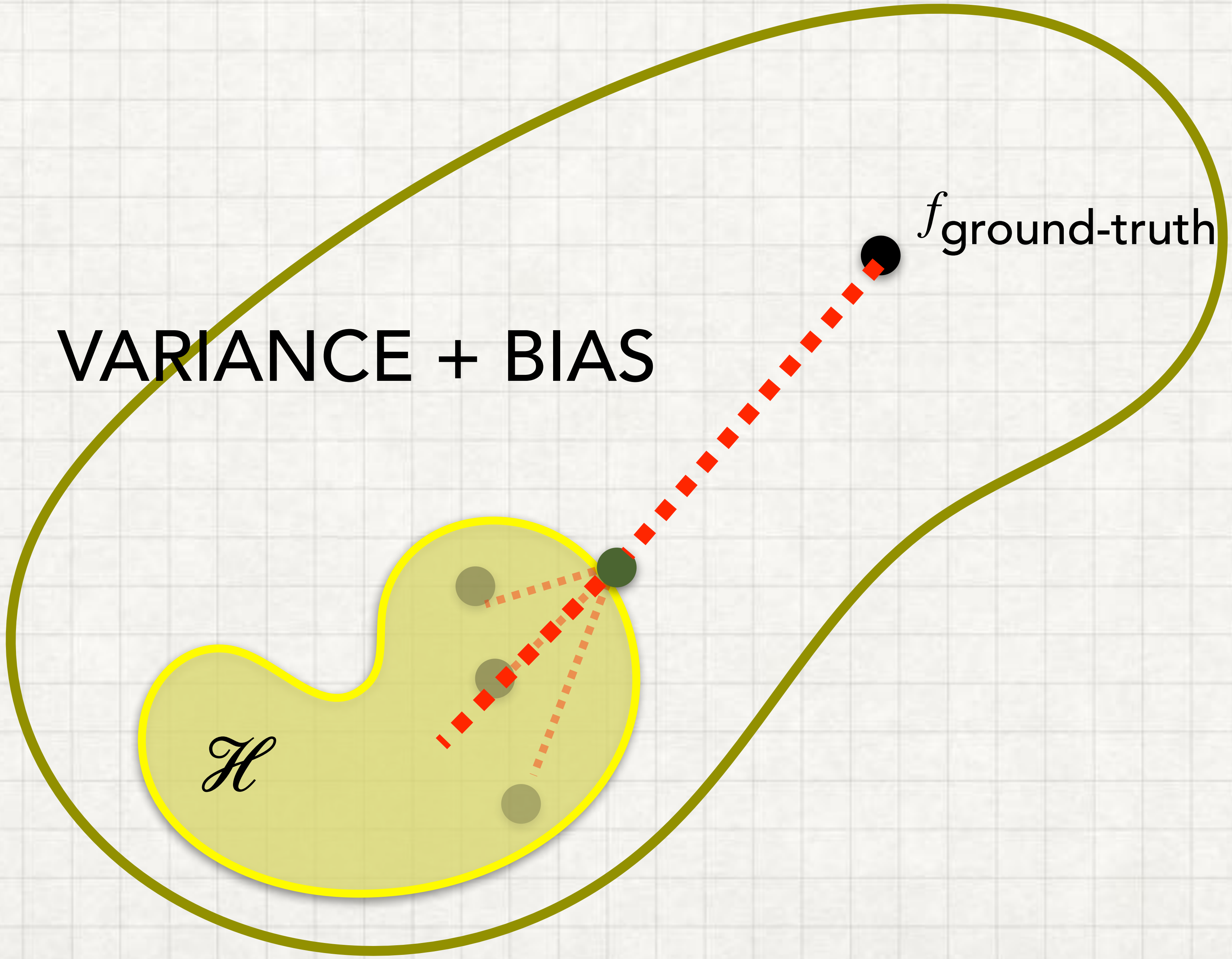
RISK OF TRAINING WITH ONE, RANDOM

D_{train}



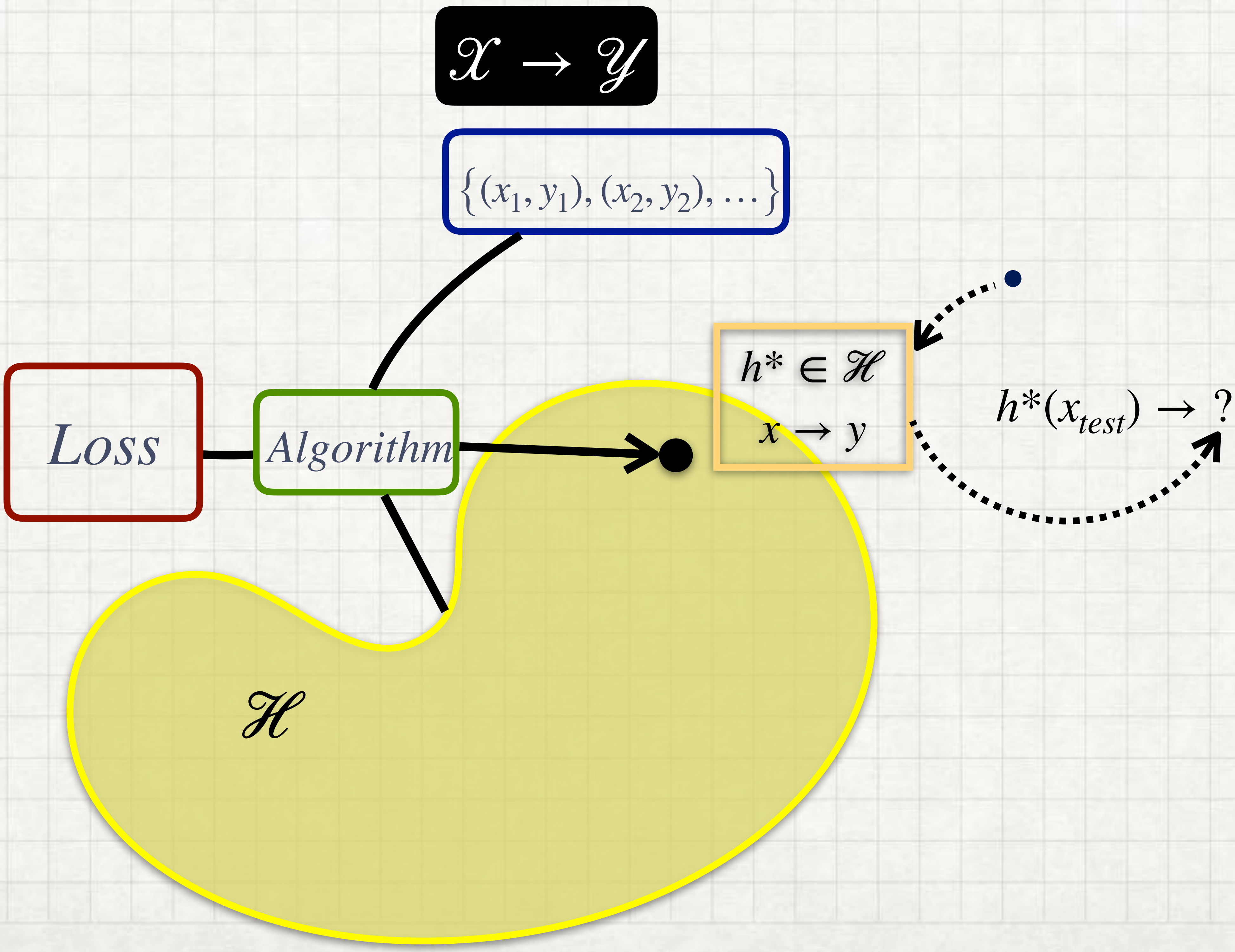
TOTAL RISK OF A TRAINING

- The expected risk of the hypothesis selected by the learning framework consists of the following two parts.



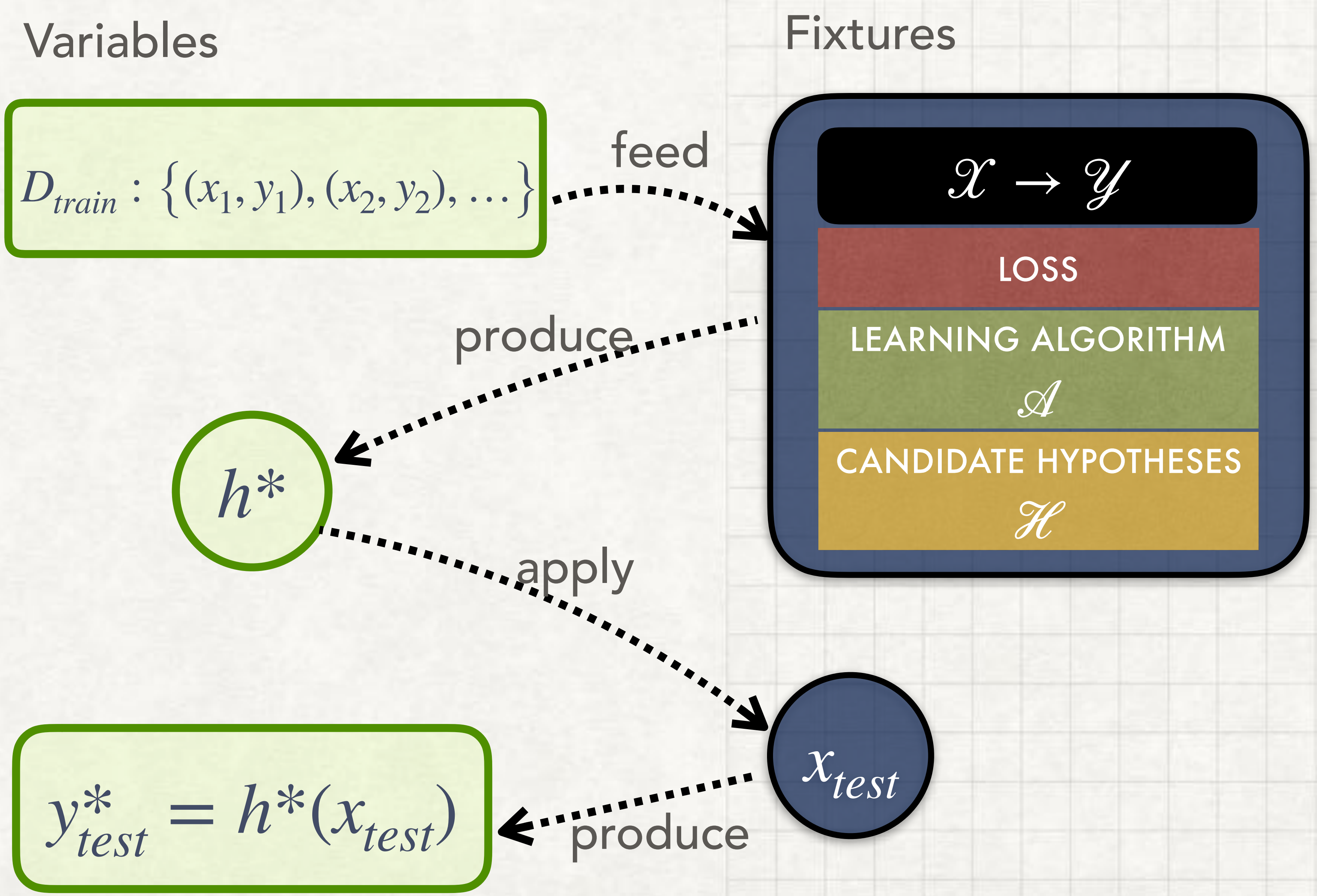
SUMMARISE: RANDOMISATION OF THE PREDICTION

- "Microscope of risk", consider the generalisation error at one x_{test}.



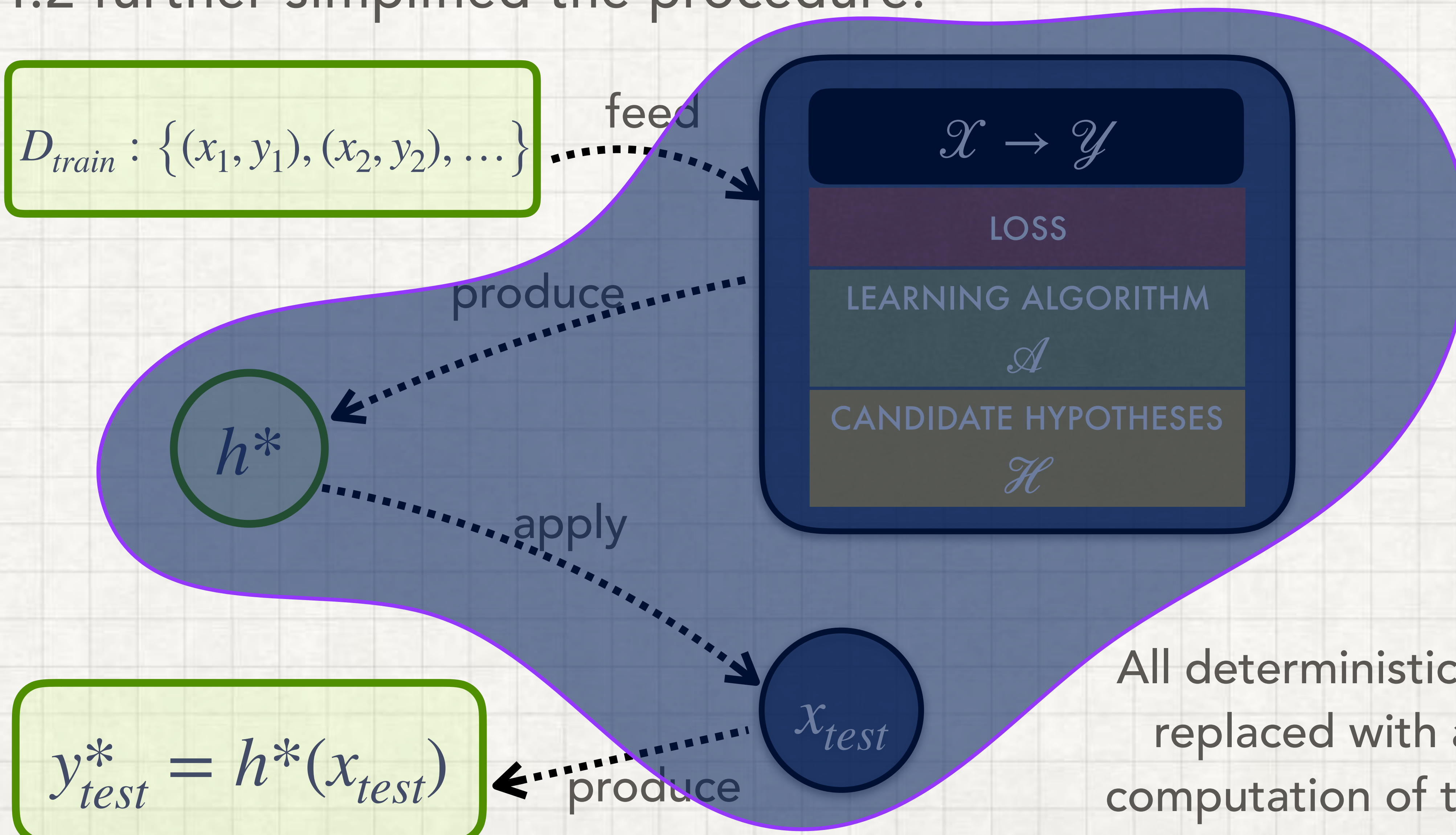
RANDOMISATION ANALYSIS

- The only factor that is inevitably random is the data sample D_{train}, so the final result of the following flow is random.



NOTEBOOK STUDY

- 4.2 and 4.3, Skip the “bootstrapping” parts (the notebook contains Python implementation, which must respect the structure of the contents, rather than the progress of ideas.)
  - 4.2 further simplified the procedure:



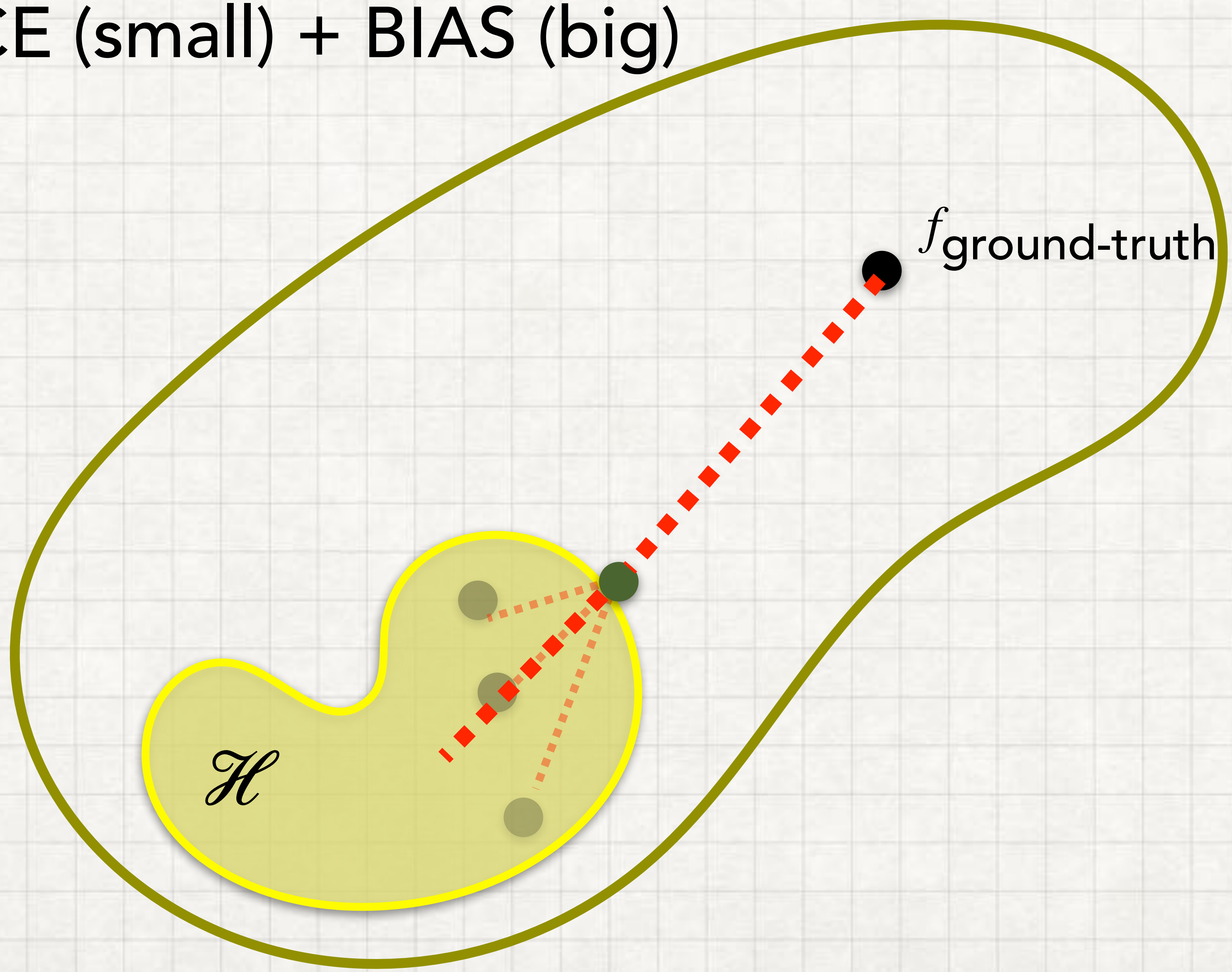
All deterministic steps are replaced with a simple computation of the sample median.

NOTEBOOK STUDY

- 4.2 and 4.3, Skip the “bootstrapping” parts (the notebook contains Python implementation, which must respect the structure of the contents, rather than the progress of ideas.)
    - 4.2 simplify -> median computation.
    - 4.3 Experiment with the first Exercise in “Bias and Variance Explained”.

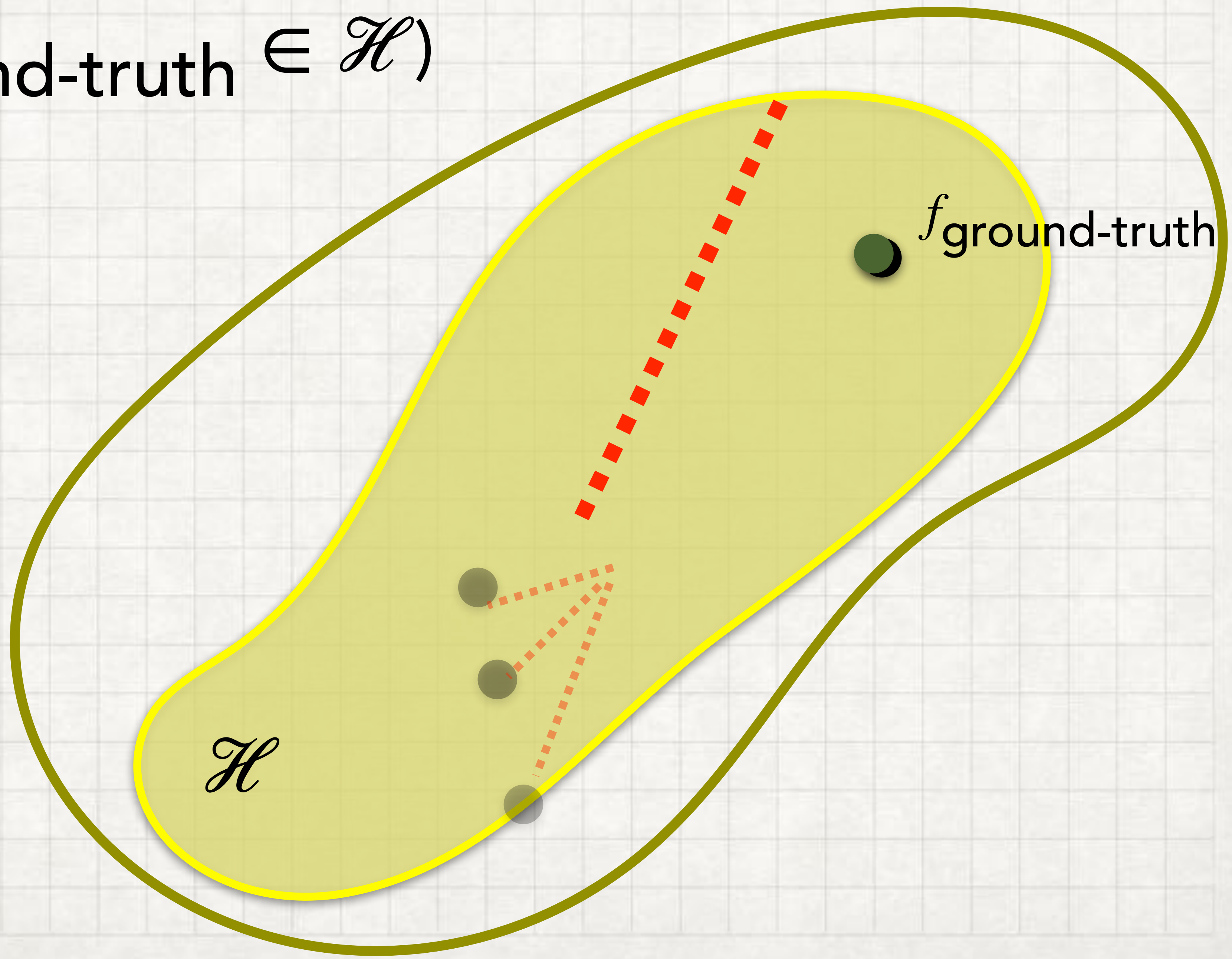
TOTAL RISK OF A TRAINING
SMALL $\mathscr{H}$

VARIANCE (small) + BIAS (big)



TOTAL RISK OF A TRAINING BIG $\mathscr{H}$

VARIANCE (big) + BIAS (small/zero — when $f_{\text{ground-truth}} \in \mathscr{H}$)



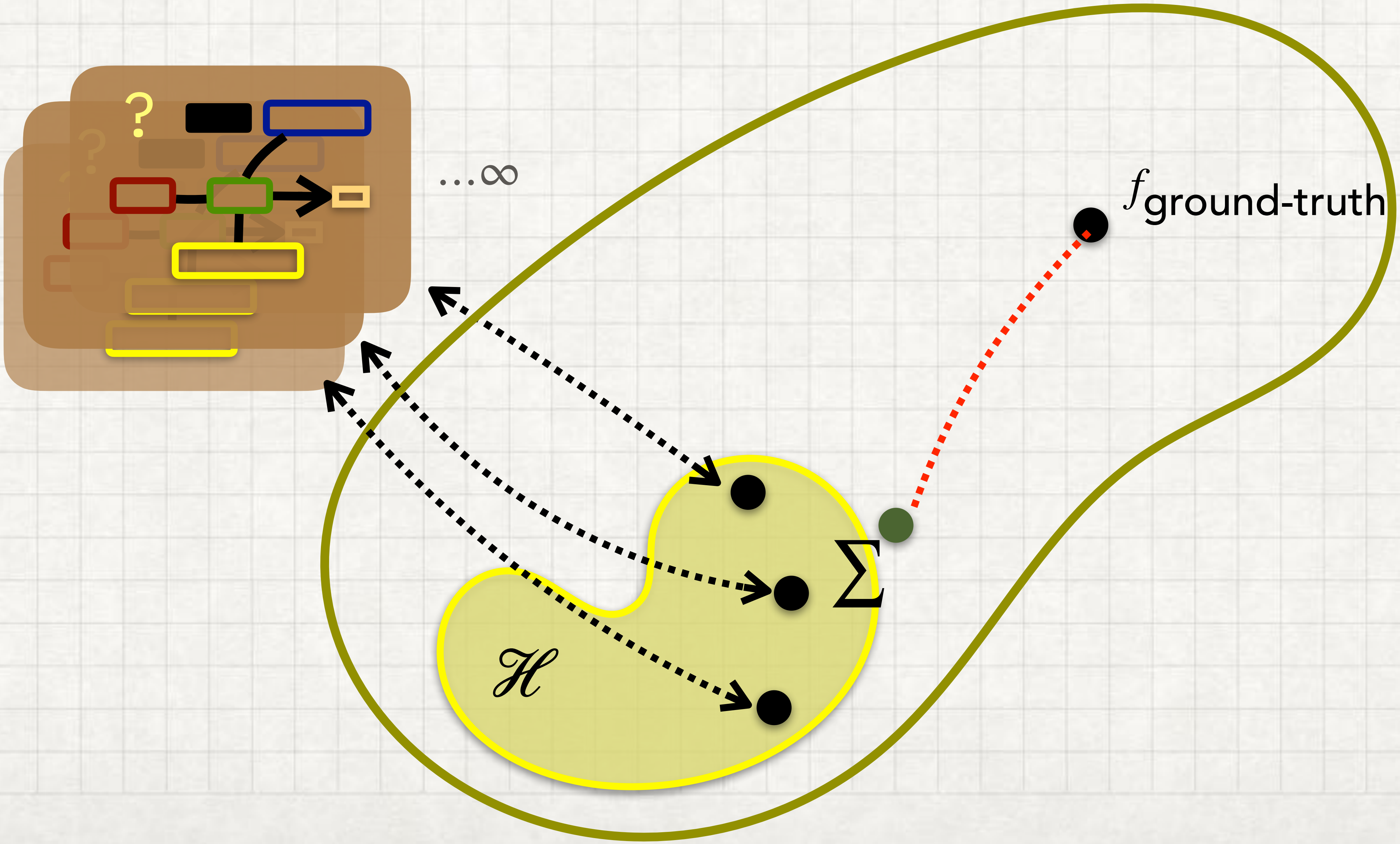
MOD3: ASSESSING AND ADDRESSING VARIANCE USING A SINGLE D_{train}

REVIEW: BOOTSTRAP / BAGGING

- IN: Data D_{train}, Model Family and Learning Method
- Repeat $bootstrap-number times: do
  - Re-sample D_{train} with replacement, get $D_{bootstrap,i}$
  - Train model, get h_i^B
- Prediction: Aggregate $h_i^B(x_{test})$'s

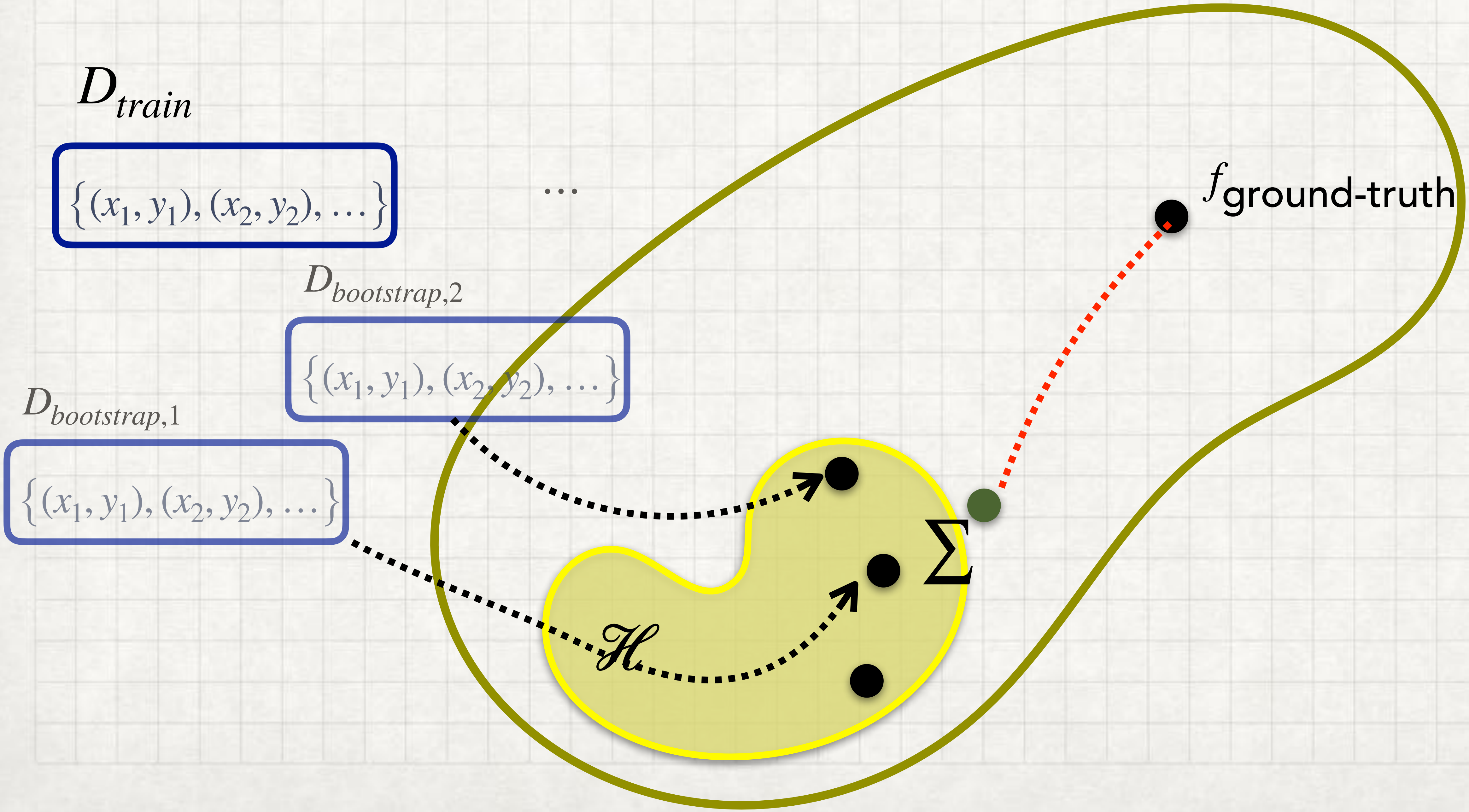
BAGGING IS AN ATTEMPT TO ACHIEVE "INFINITE" MODEL AVERAGING

- The bootstrap samples are used to assess statistical information and reduce the variance due to the randomisation of D_{train}



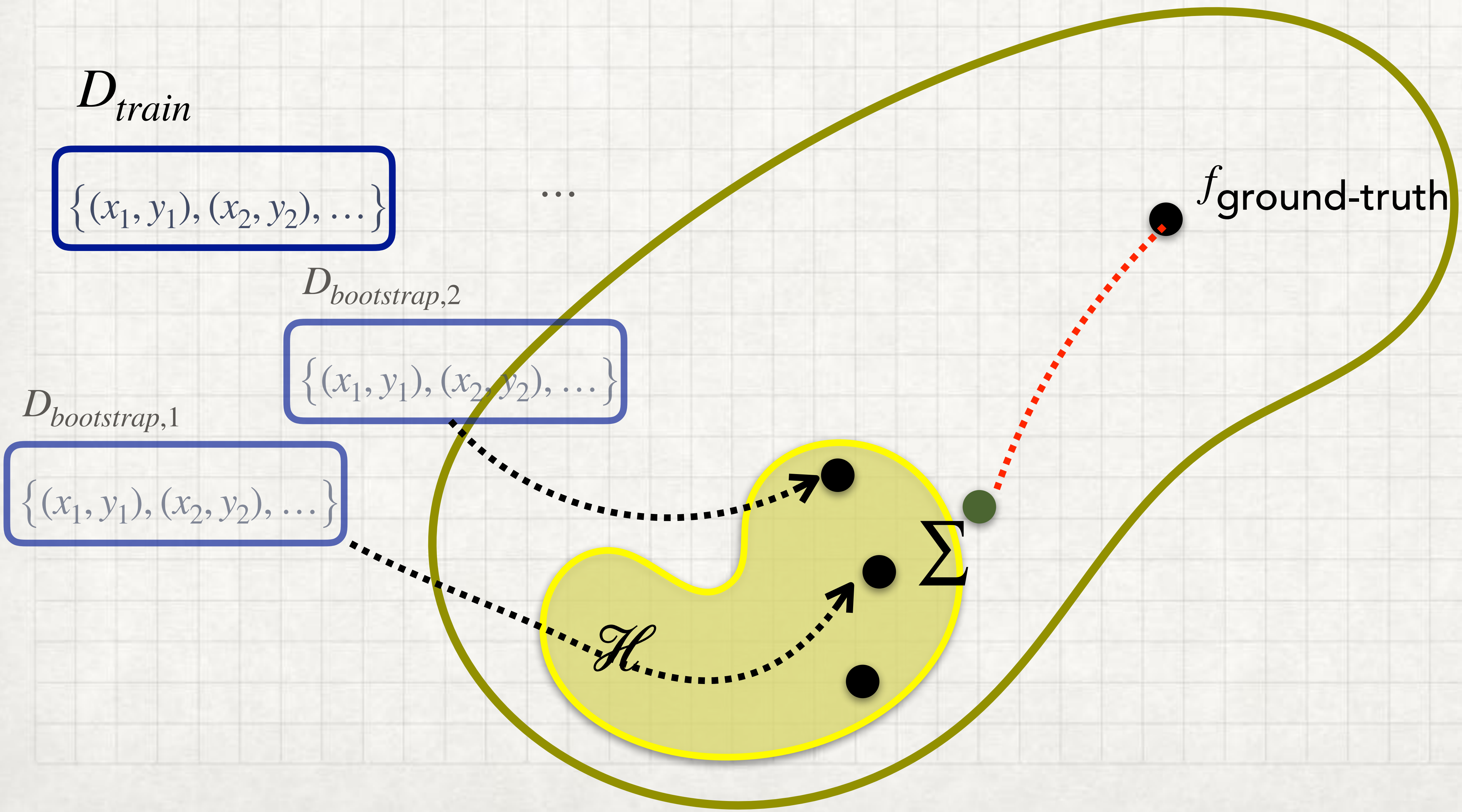
BOOTSTRAP SAMPLES

- The bootstrap samples are used to assess statistical information and reduce the variance due to the randomisation of D_{train}



BOOTSTRAP: FREE LUNCH?

- Bootstrap cannot change the bias. But it can reduce variance given right settings at the cost of computation.



NOTEBOOK STUDY

- 4.2 “bootstrap” parts
  - What is bootstrap samples.
  - Experiment and contemplate: For the estimation of median from samples, how bootstrap (resampling) helps reduce variance.

- 4.3 Observe Bias and Variance in sample model (if not done in the previous notebook study).

- 4.4 Bootstrap helps reduce variance in some settings.

THANKS