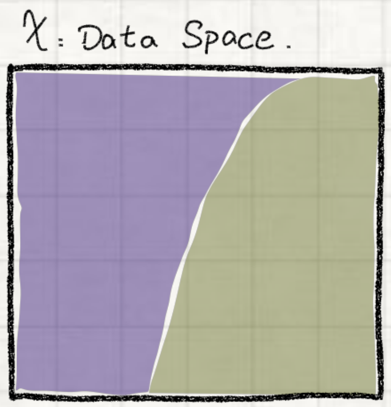
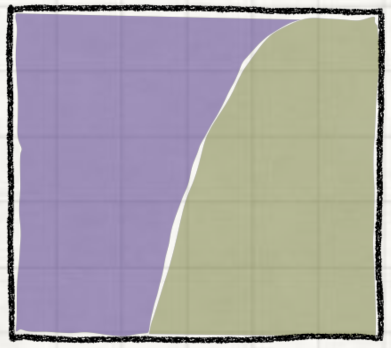
BIAS VARIANCE BOOTSTRAPPING AND BOOSTING

REVIEW

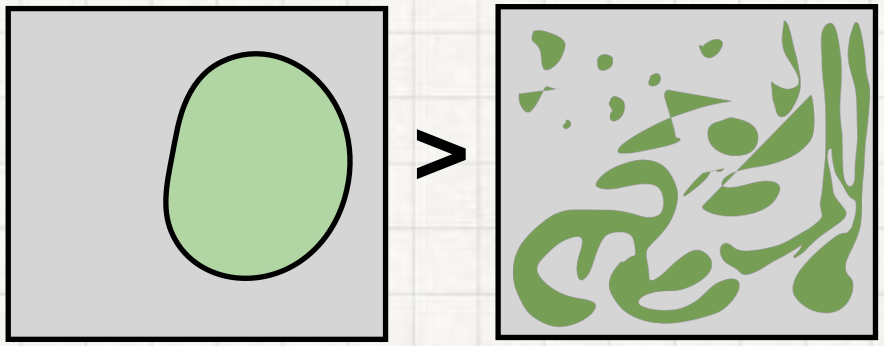
REVIEW



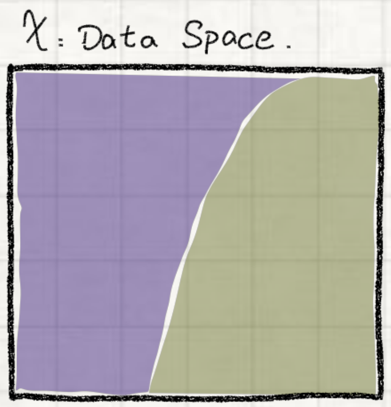
REVIEW



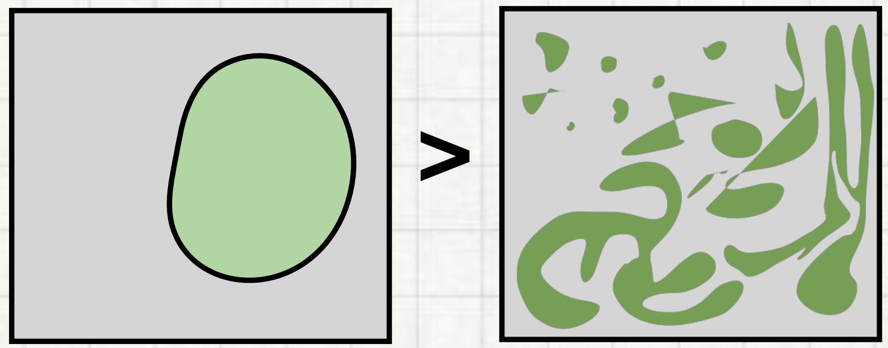
- 2^{2500}=
3758280234548012036833624189
7238650486773655175925867705
6523839782231681498337708535
7327257526588443337024577495
2605776030922789135161776565
1907310968780236464694043316
2365621467244164785911318325
9372911122158018053174923277
7515579969899075142213969117
9948773438020494216249544022
1452939078164756333953502477
2584901607666862982567918622
8496361602088773658349501637
9018852302624744050739038203
2188892386109905869706753143
2439211984822120754440224333
6655478685655938968958563812
6582377224037721702239991441
4660261857526515029364722809
1101850032037549633674995156
9521541850441747925844066295
2796718726052857925526601307
0204799821833474935632167746
9529682551765858267502715894
0078877272500707803502629523
7721402884229748626359787979
2176338220932619489509376

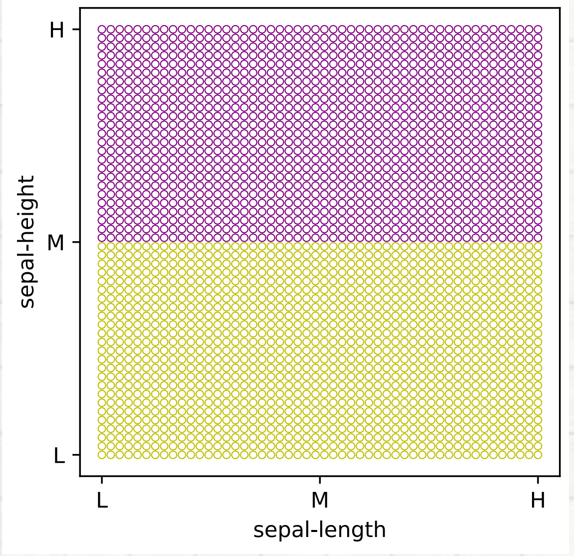


REVIEW



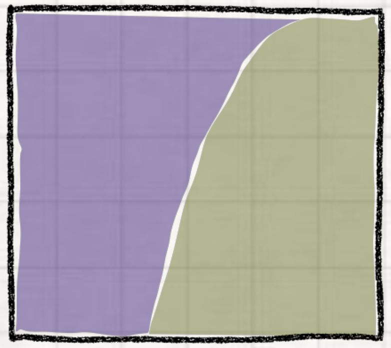
- $2^{2500}=$
  375828023454801203683362418972386504867736551759258677056523839782231681498337708535732725752658844333702457749526057760309227891351617765651907310968780236464694043316236562146724416478591131832593729111221580180531749232777515579969899075142213969117994877343802049421624954402214529390781647563339535024772584901607666862982567918622849636160208877365834950163790188523026247440507390382032188892386109905869706753143243921198482212075444022433366554786856559389689585638126582377224037721702239991441466026185752651502936472280911018500320375496336749951569521541850441747925844066295279671872605285792552660130702047998218334749356321677469529682551765858267502715894007887727250070780350262952377214028842297486263597879792176338220932619489509376



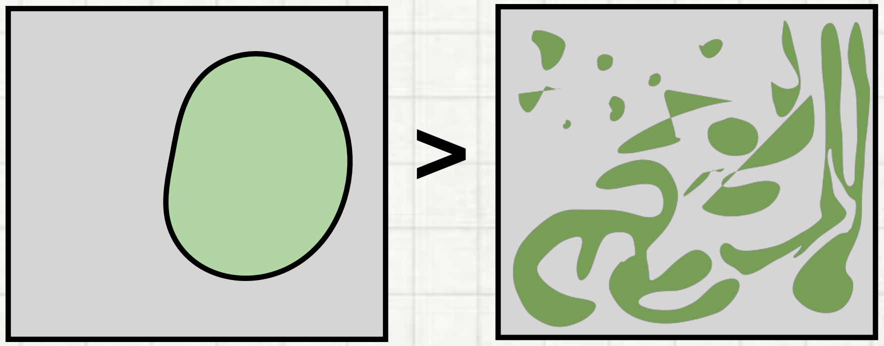


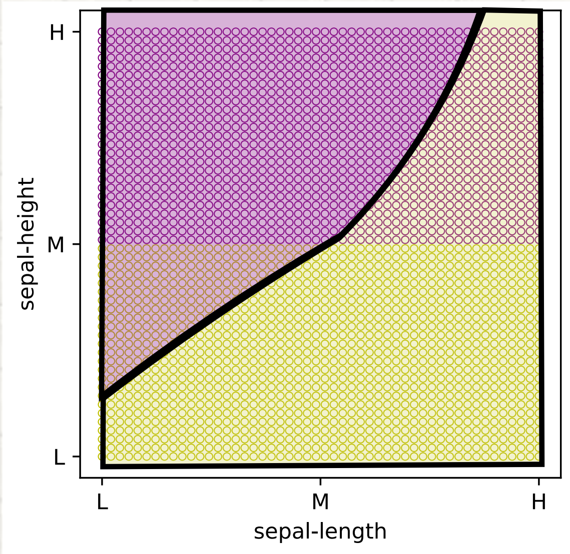
REVIEW

χ: Data Space.

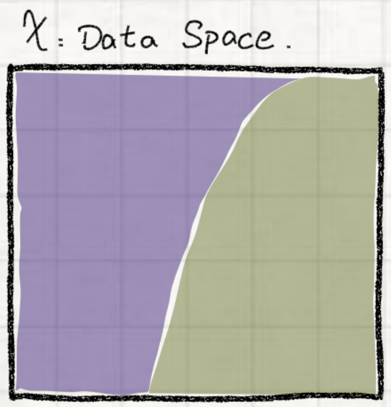


- $2^{2500}=$
3758280234548012036833624189
7238650486773655175925867705
6523839782231681498337708535
7327257526588443337024577495
2605776030922789135161776565
1907310968780236464694043316
2365621467244164785911318325
9372911122158018053174923277
7515579969899075142213969117
9948773438020494216249544022
1452939078164756333953502477
2584901607666862982567918622
8496361602088773658349501637
9018852302624744050739038203
2188892386109905869706753143
2439211984822120754440224333
6655478685655938968958563812
6582377224037721702239991441
4660261857526515029364722809
1101850032037549633674995156
9521541850441747925844066295
2796718726052857925526601307
0204799821833474935632167746
9529682551765858267502715894
0078877272500707803502629523
7721402884229748626359787979
2176338220932619489509376

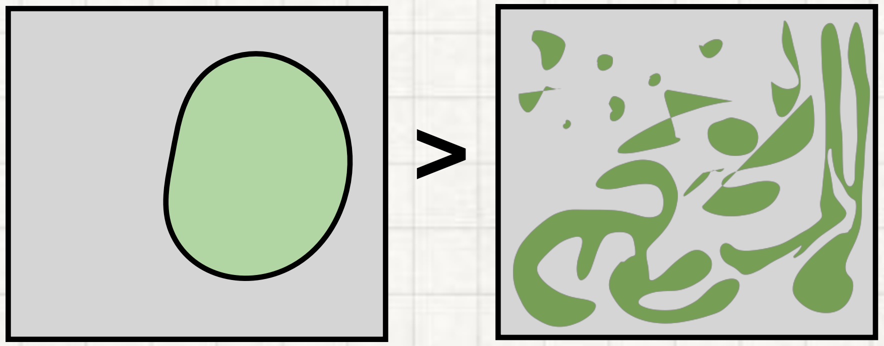


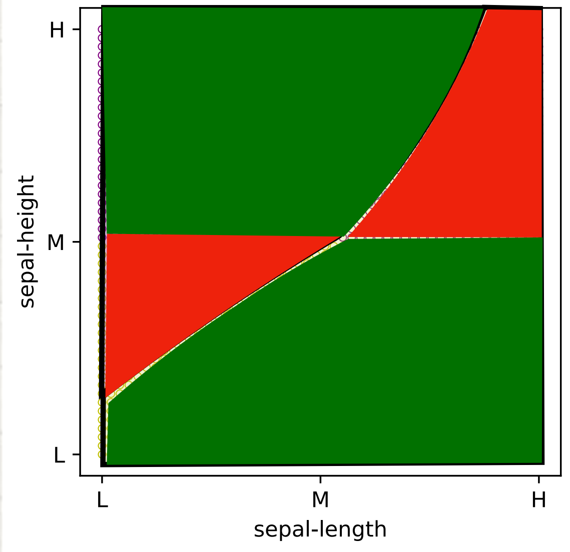


REVIEW



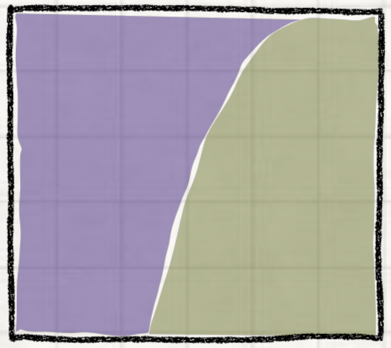
- $2^{2500}=$ 375828023454801203683362418972386504867736551759258677056523839782231681498337708535732725752658844333702457749526057760309227891351617765651907310968780236464694043316236562146724416478591131832593729111221580180531749232777515579969899075142213969117994877343802049421624954402214529390781647563339535024772584901607666862982567918622849636160208877365834950163790188523026247440507390382032188892386109905869706753143243921198482212075444022433366554786856559389689585638126582377224037721702239991441466026185752651502936472280911018500320375496336749951569521541850441747925844066295279671872605285792552660130702047998218334749356321677469529682551765858267502715894007887727250070780350262952377214028842297486263597879792176338220932619489509376



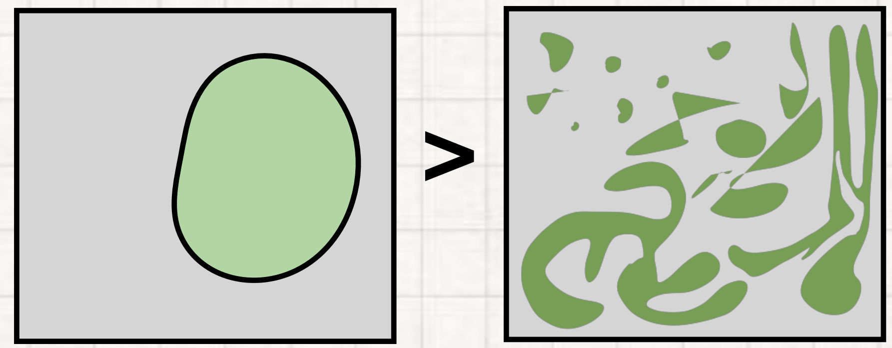


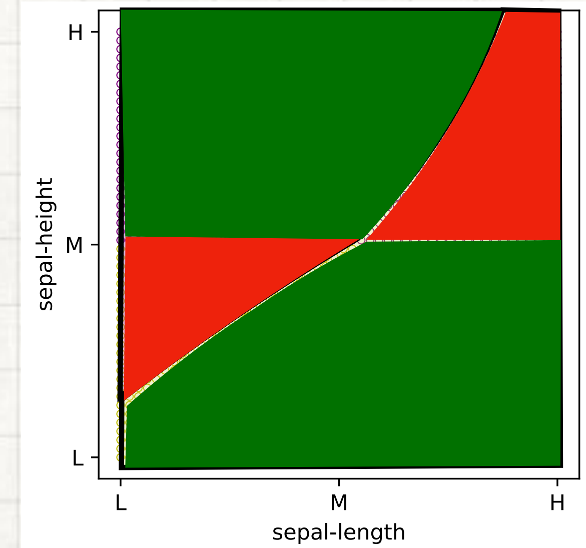
REVIEW

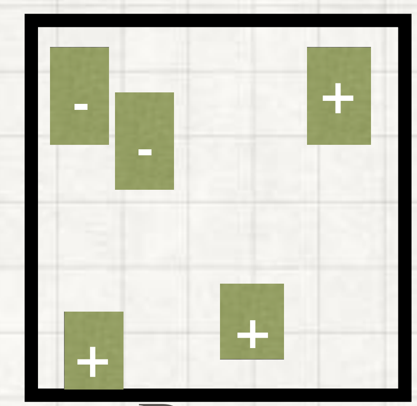
χ: Data Space.



- $2^{2500}=$
3758280234548012036833624189
7238650486773655175925867705
6523839782231681498337708535
7327257526588443337024577495
2605776030922789135161776565
1907310968780236464694043316
2365621467244164785911318325
9372911122158018053174923277
7515579969899075142213969117
9948773438020494216249544022
1452939078164756333953502477
2584901607666862982567918622
8496361602088773658349501637
9018852302624744050739038203
2188892386109905869706753143
2439211984822120754440224333
6655478685655938968958563812
6582377224037721702239991441
4660261857526515029364722809
1101850032037549633674995156
9521541850441747925844066295
2796718726052857925526601307
0204799821833474935632167746
9529682551765858267502715894
0078877272500707803502629523
7721402884229748626359787979
2176338220932619489509376

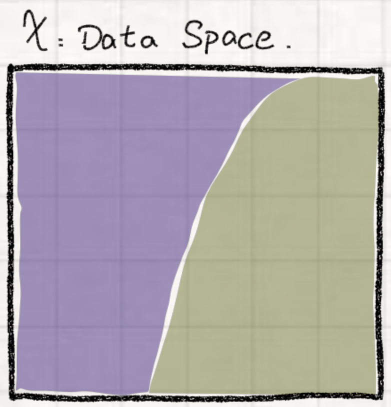




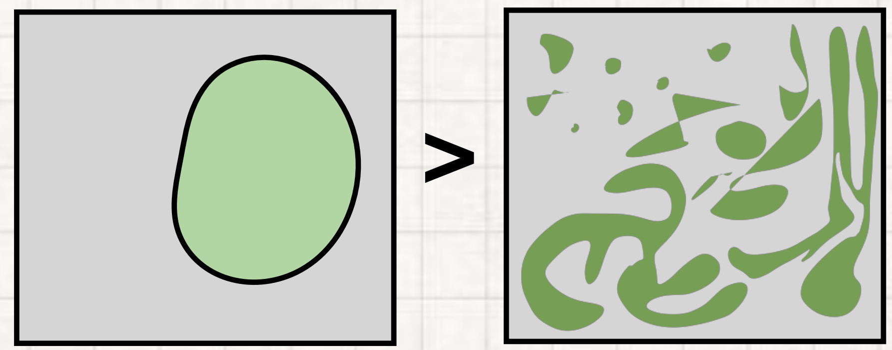


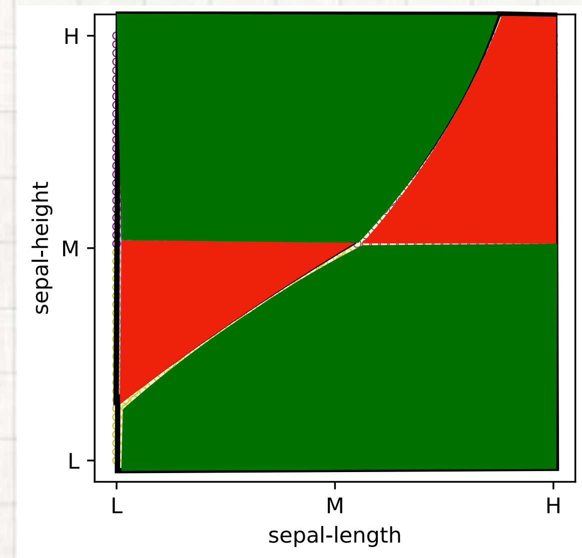
D_{train}

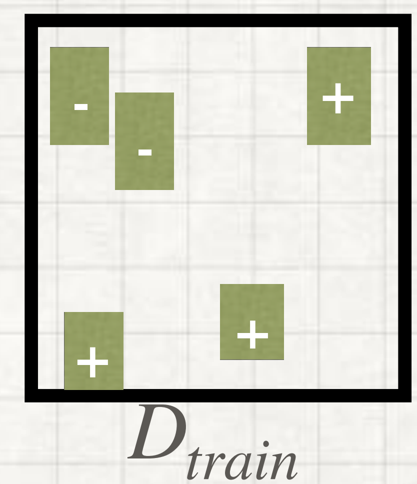
REVIEW

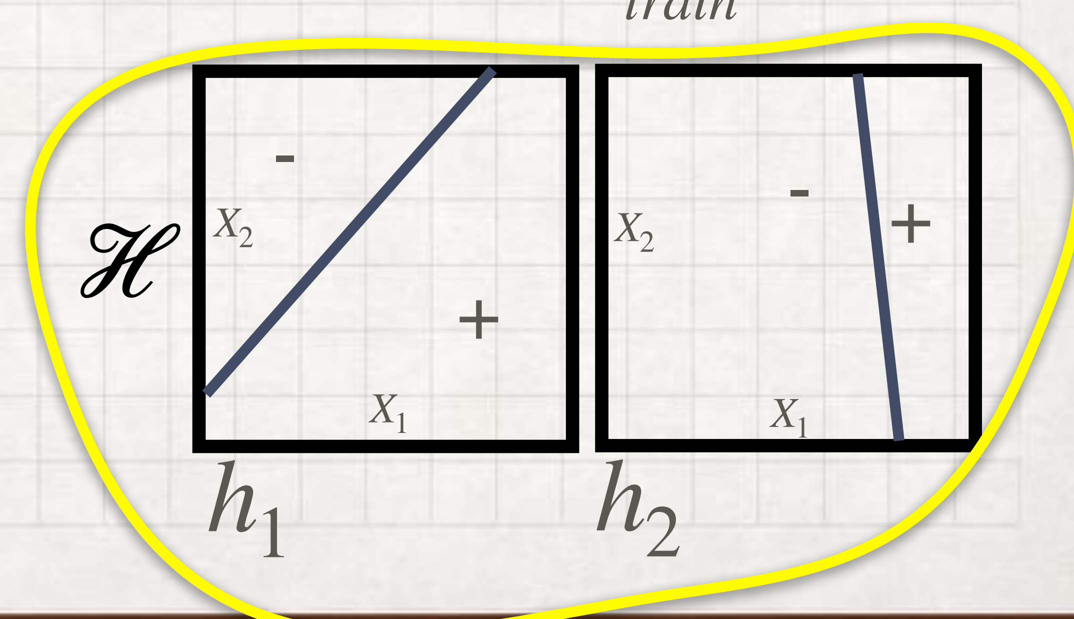


2^{2500}=
3758280234548012036833624189
7238650486773655175925867705
6523839782231681498337708535
7327257526588443337024577495
2605776030922789135161776565
1907310968780236464694043316
2365621467244164785911318325
9372911122158018053174923277
7515579969899075142213969117
9948773438020494216249544022
1452939078164756333953502477
2584901607666862982567918622
8496361602088773658349501637
9018852302624744050739038203
2188892386109905869706753143
2439211984822120754440224333
6655478685655938968958563812
6582377224037721702239991441
4660261857526515029364722809
1101850032037549633674995156
9521541850441747925844066295
2796718726052857925526601307
0204799821833474935632167746
9529682551765858267502715894
0078877272500707803502629523
7721402884229748626359787979
2176338220932619489509376



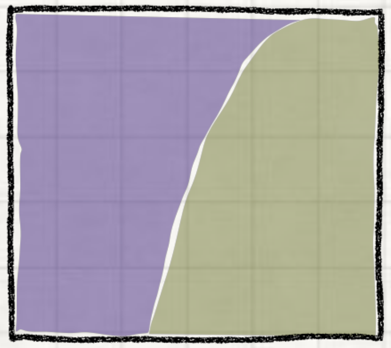




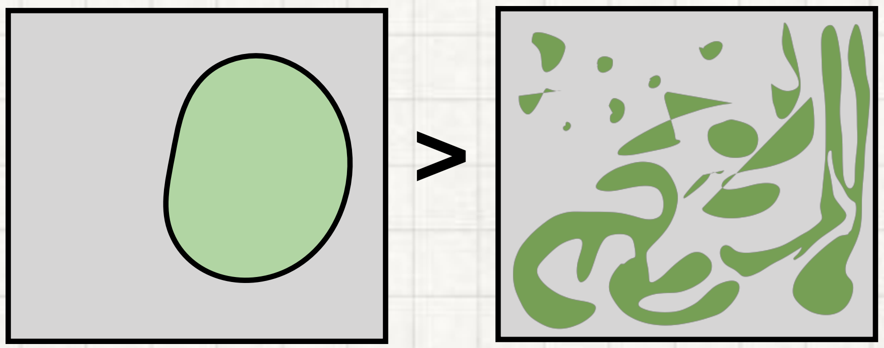


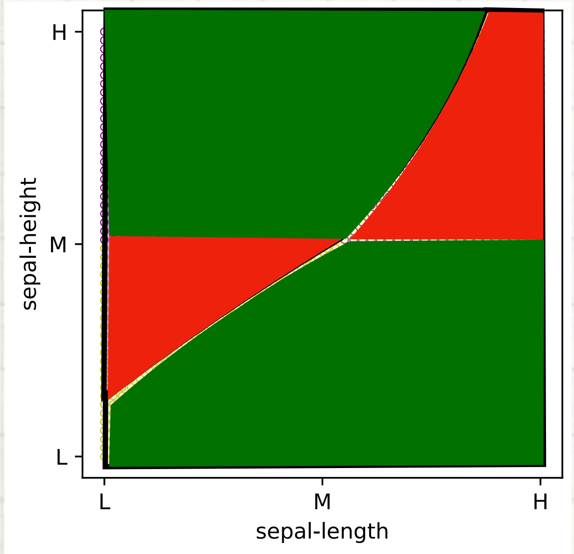
REVIEW

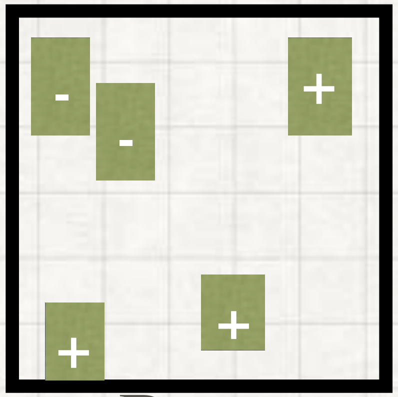
χ: Data Space.



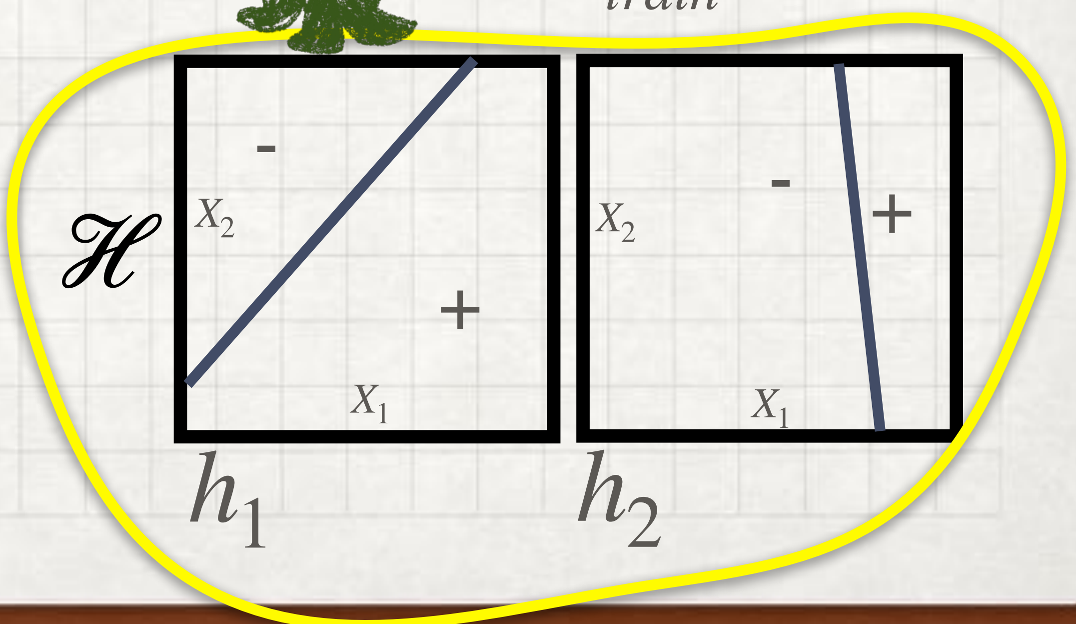
* 2^{2500}=
375828023454801203683362418972386504867736551759258677056523839782231681498337708535732725752658844333702457749526057760309227891351617765651907310968780236464694043316236562146724416478591131832593729111221580180531749232777515579969899075142213969117994877343802049421624954402214529390781647563339535024772584901607666862982567918622849636160208877365834950163790188523026247440507390382032188892386109905869706753143243921198482212075444022433366554786856559389689585638126582377224037721702239991441466026185752651502936472280911018500320375496336749951569521541850441747925844066295279671872605285792552660130702047998218334749356321677469529682551765858267502715894007887727250070780350262952377214028842297486263597879792176338220932619489509376



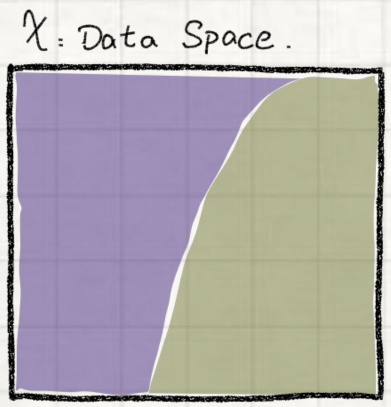


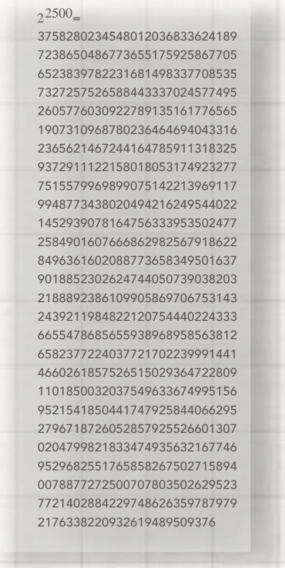


D_{train}



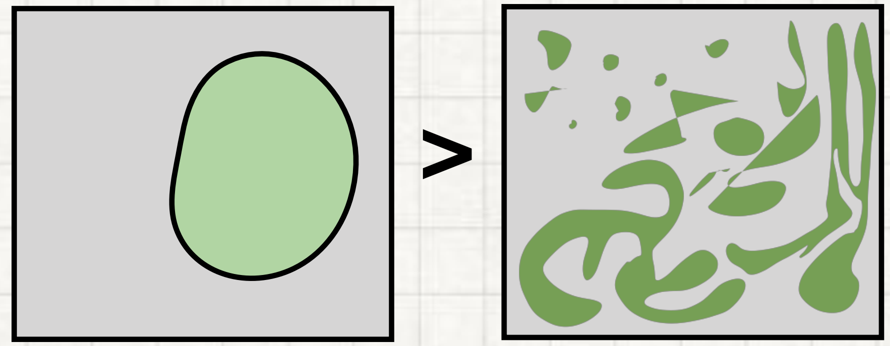
REVIEW

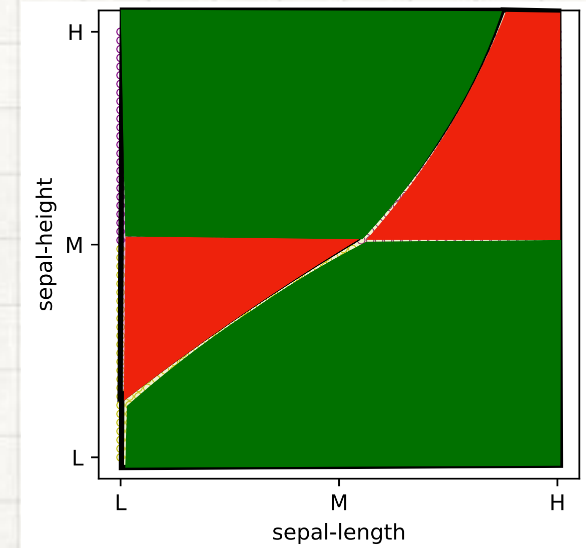


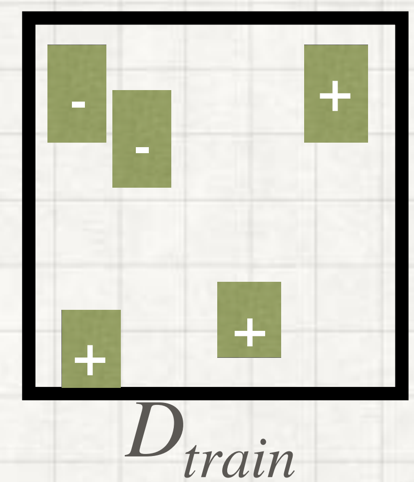


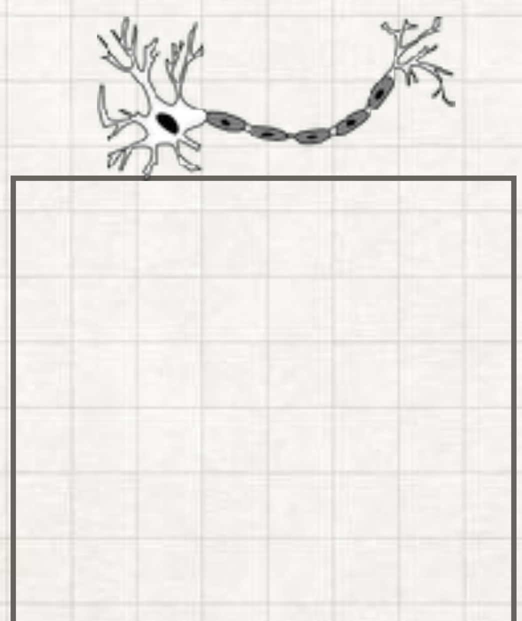
$2^{2500}=$

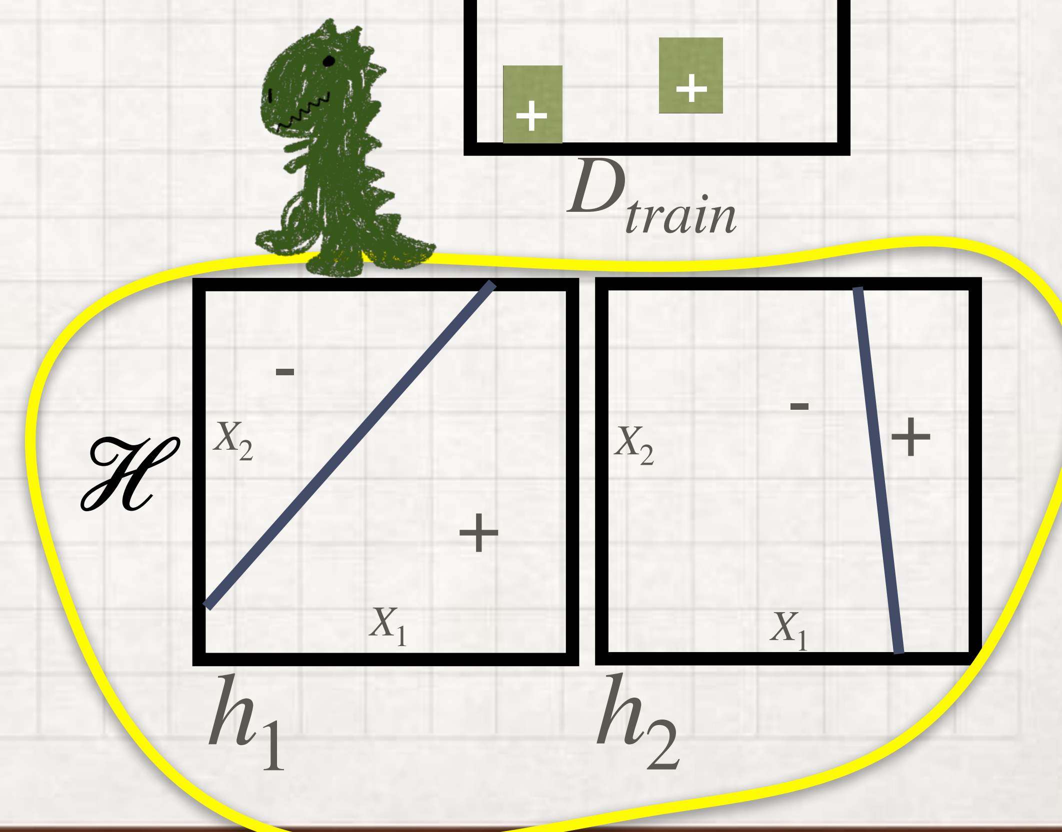
3758280234548012036833624189
7238650486773655175925867705
6523839782231681498337708535
7327257526588443337024577495
2605776030922789135161776565
1907310968780236464694043316
2365621467244164785911318325
9372911122158018053174923277
7515579969899075142213969117
9948773438020494216249544022
1452939078164756333953502477
2584901607666862982567918622
8496361602088773658349501637
9018852302624744050739038203
2188892386109905869706753143
2439211984822120754440224333
6655478685655938968958563812
6582377224037721702239991441
4660261857526515029364722809
1101850032037549633674995156
9521541850441747925844066295
2796718726052857925526601307
0204799821833474935632167746
9529682551765858267502715894
0078877272500707803502629523
7721402884229748626359787979
217633822093261948950937 6





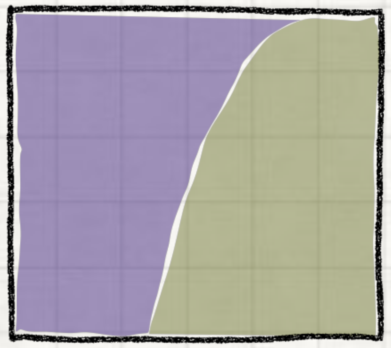


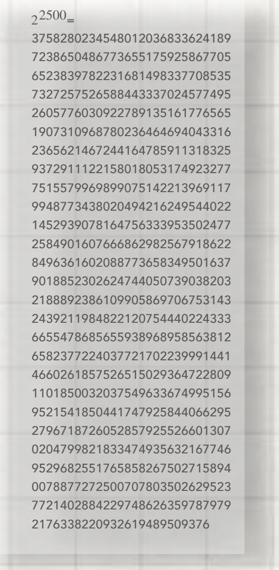


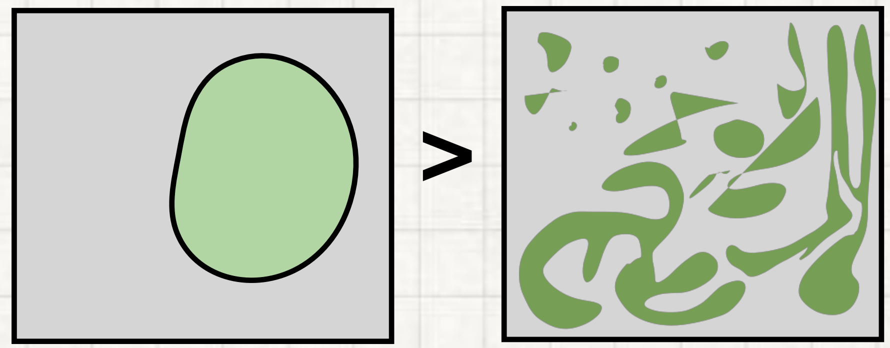


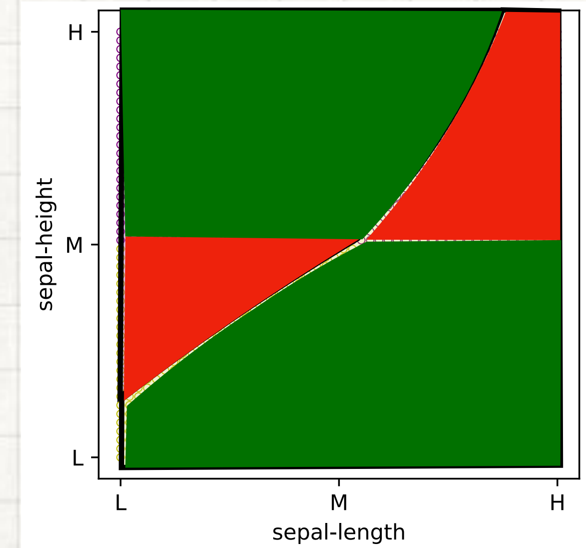
REVIEW

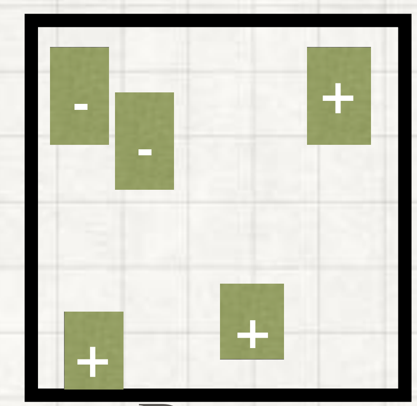
χ: Data Space.



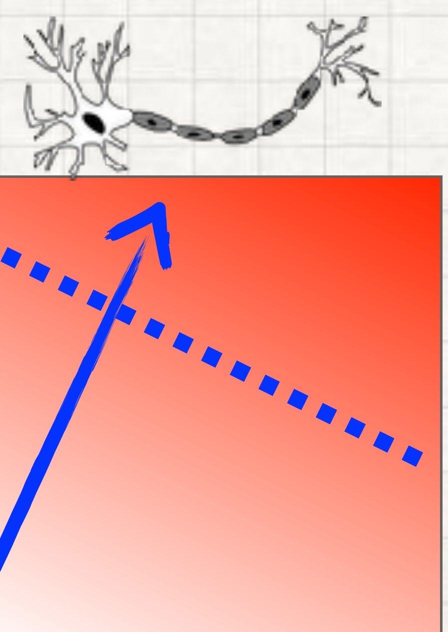


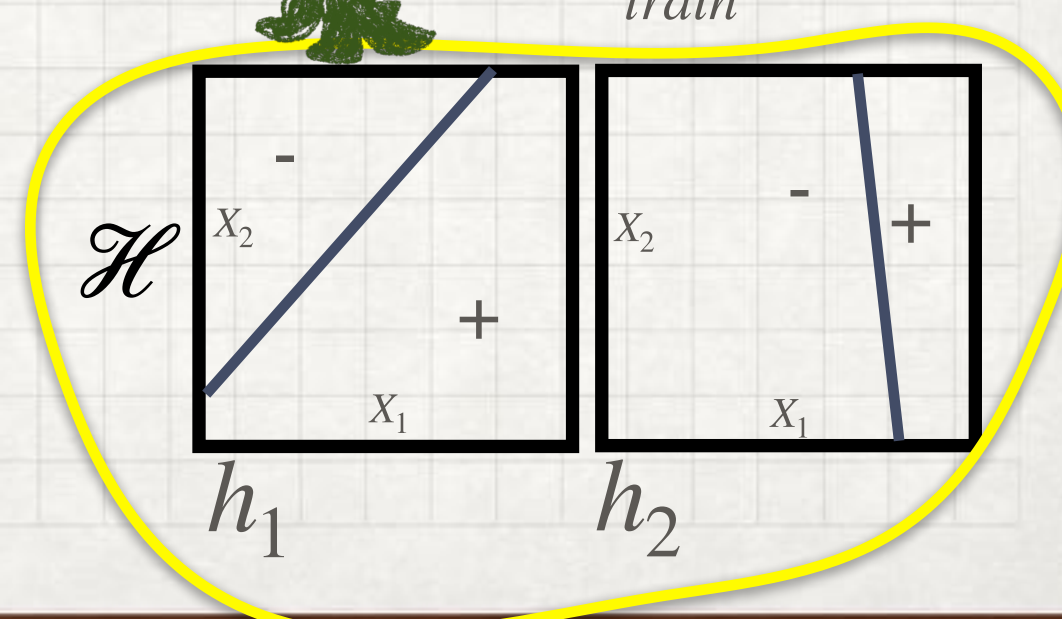






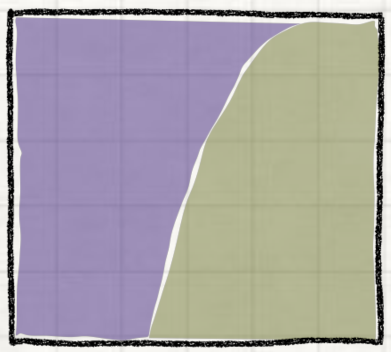
D_{train}



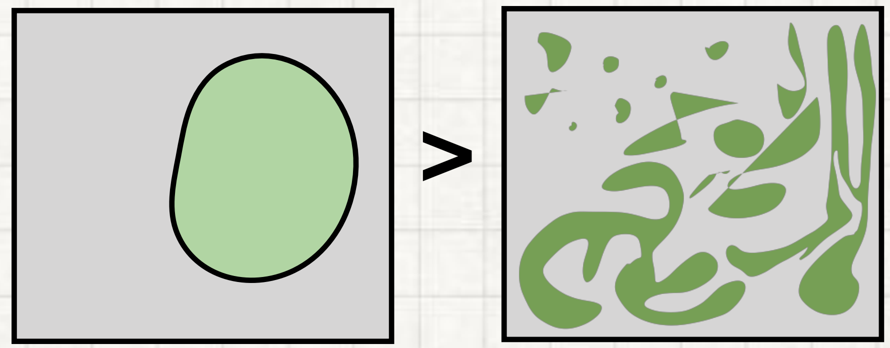


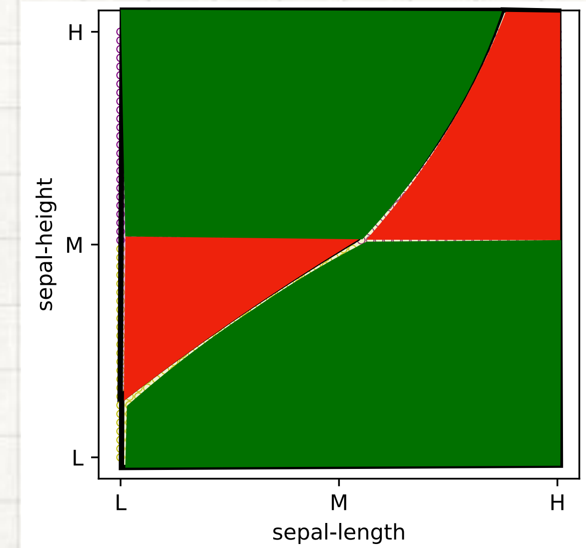
REVIEW

$\mathcal{X}$: Data Space.



2^{2500}=
3758280234548012036833624189
7238650486773655175925867705
6523839782231681498337708535
7327257526588443337024577495
2605776030922789135161776565
1907310968780236464694043316
2365621467244164785911318325
9372911122158018053174923277
7515579969899075142213969117
9948773438020494216249544022
1452939078164756333953502477
2584901607666862982567918622
8496361602088773658349501637
9018852302624744050739038203
2188892386109905869706753143
2439211984822120754440224333
6655478685655938968958563812
6582377224037721702239991441
4660261857526515029364722809
1101850032037549633674995156
9521541850441747925844066295
2796718726052857925526601307
0204799821833474935632167746
9529682551765858267502715894
0078877272500707803502629523
7721402884229748626359787979
21763382209326194895 09376





$\mathcal{X} \to \mathcal{Y}$

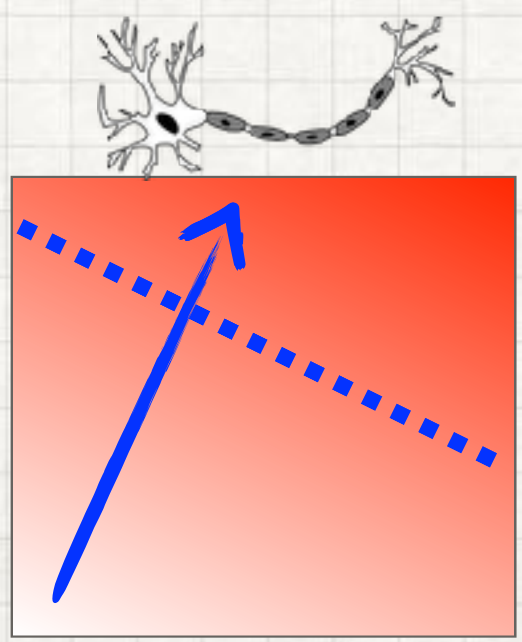
$\{(x_1, y_1), (x_2, y_2), \dots\}$

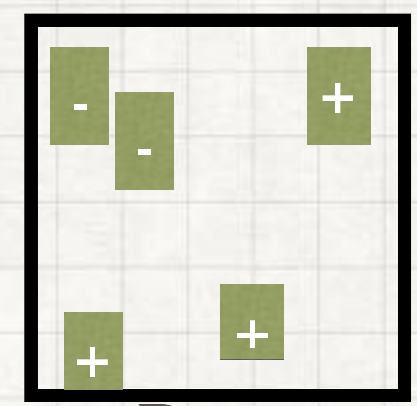
Loss

Algorithm

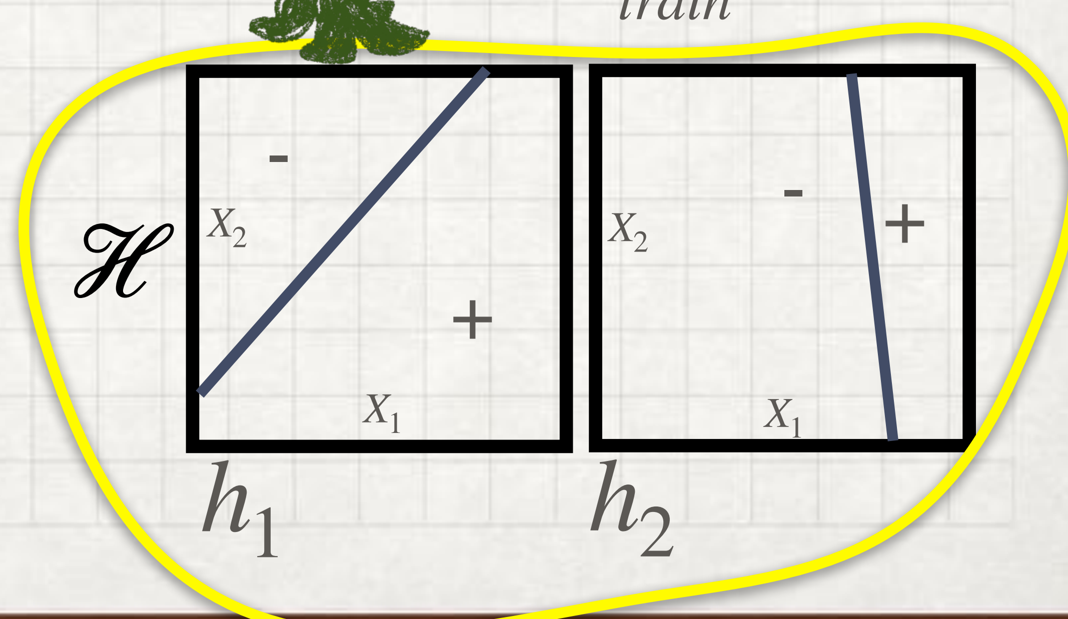
$h^* \in \mathcal{H}$

$\mathcal{H} : \{h_1, h_2, \dots, \} : x \to y$



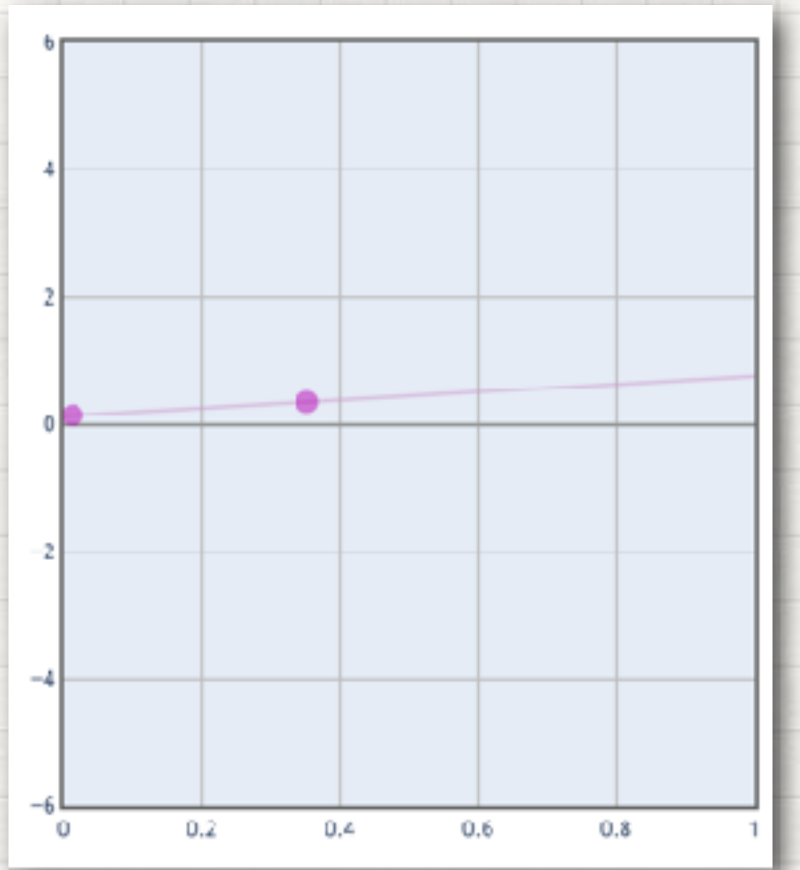


D_{train}

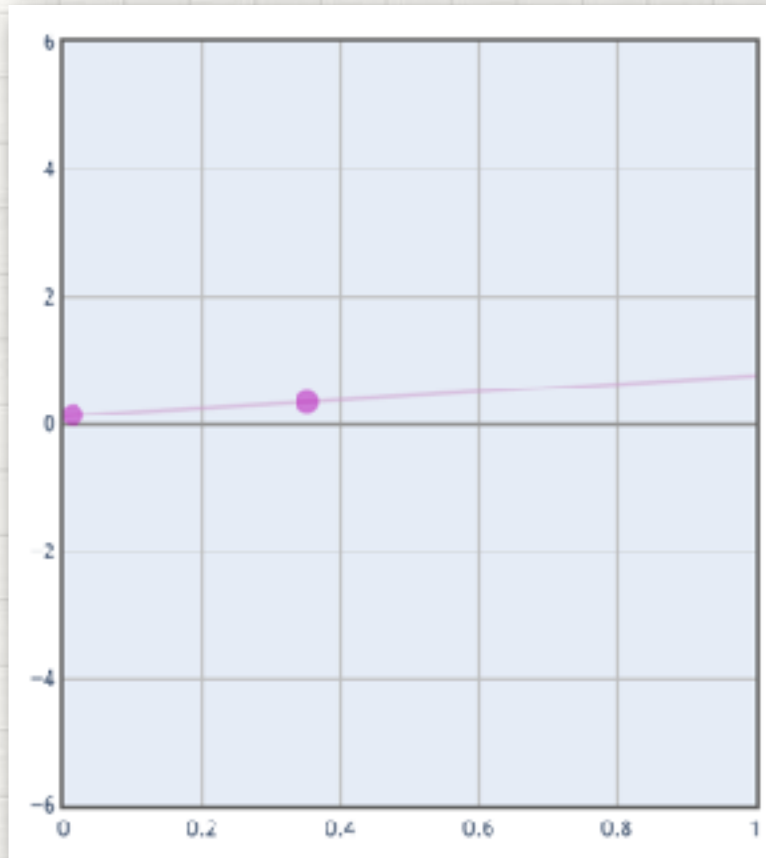


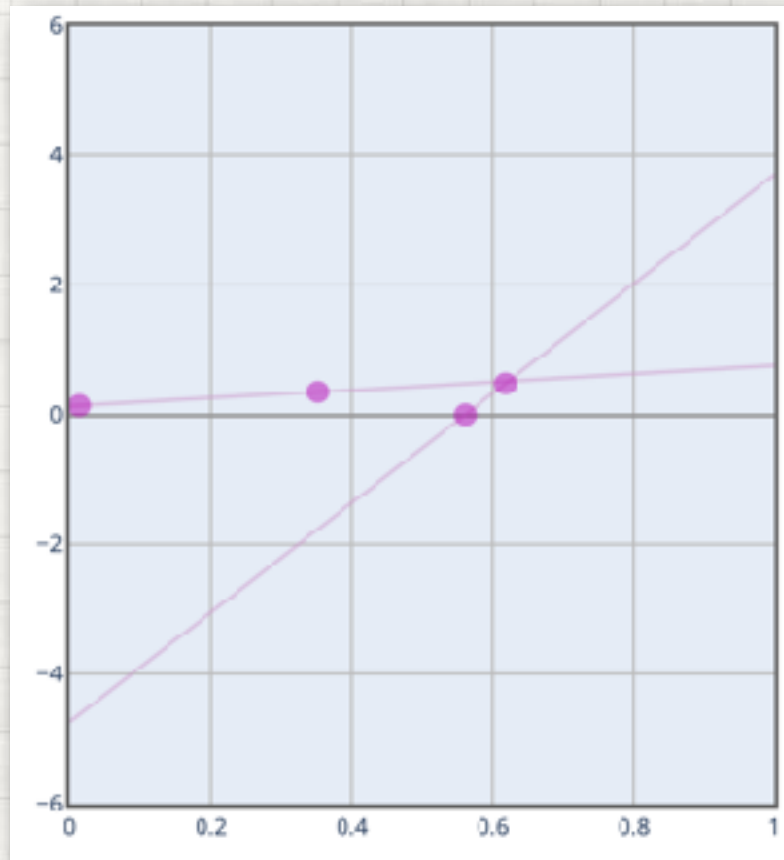
UNCERTAINTY IN LEARNING

UNCERTAINTY IN LEARNING

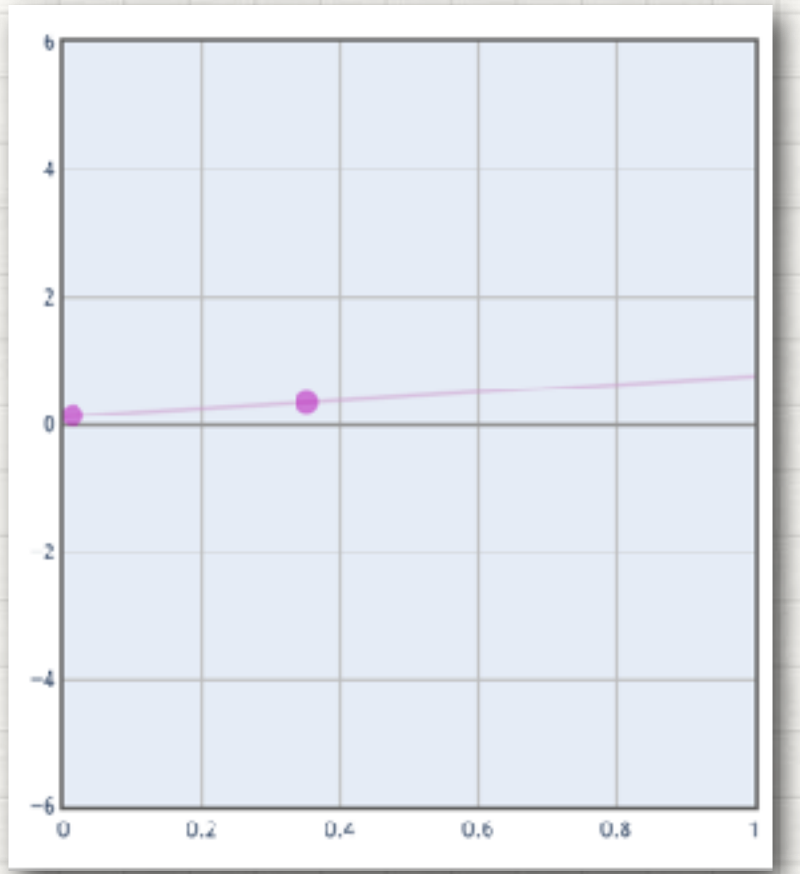


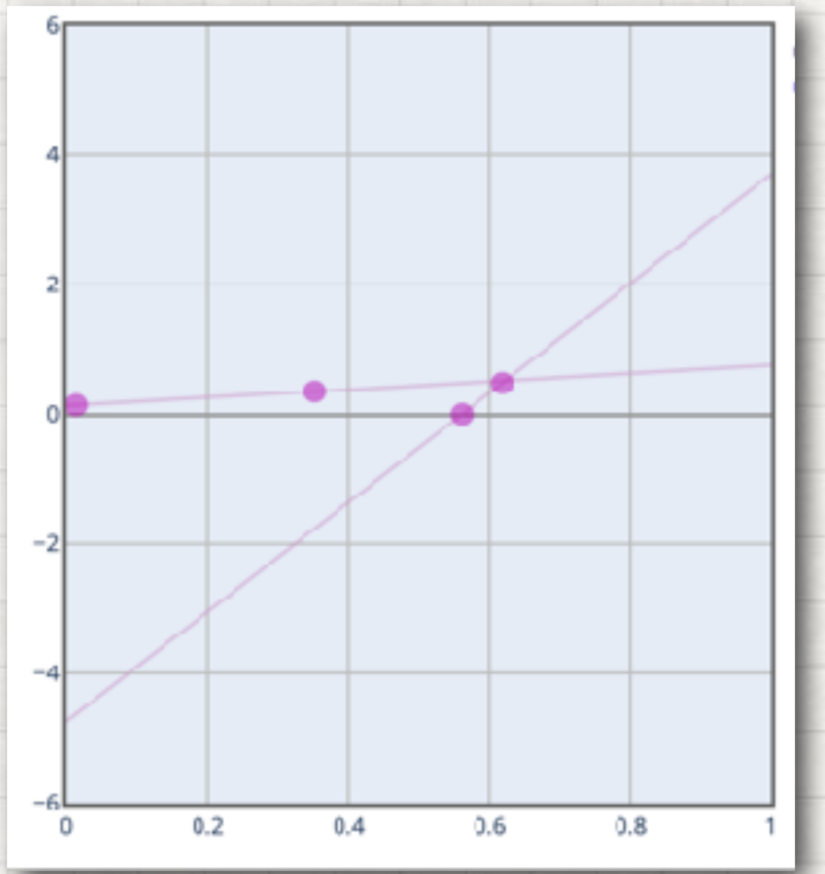
UNCERTAINTY IN LEARNING

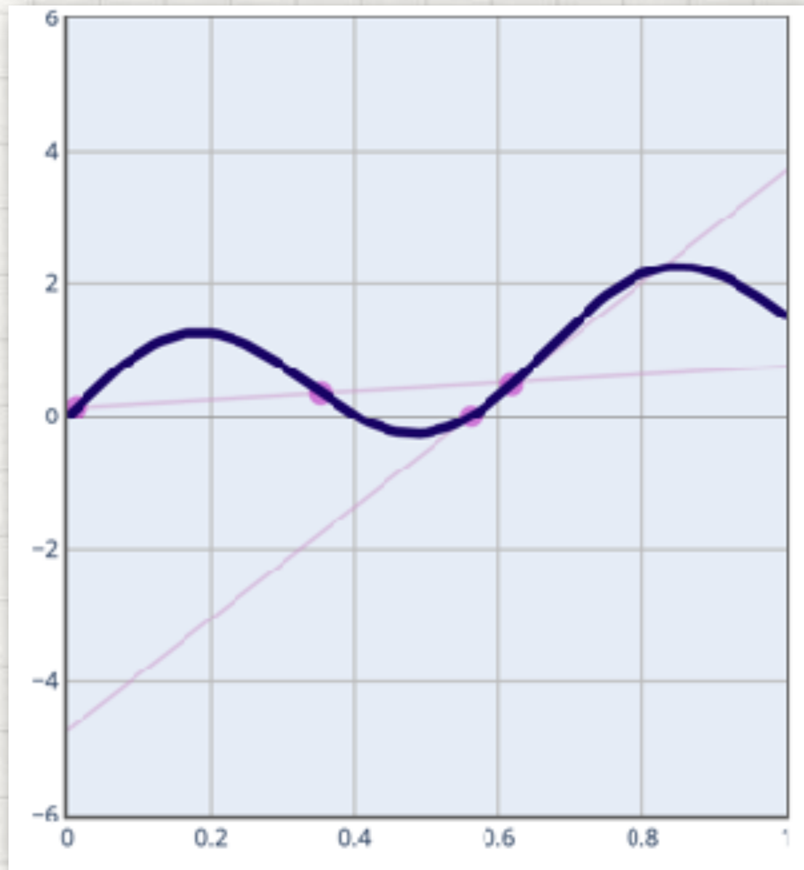




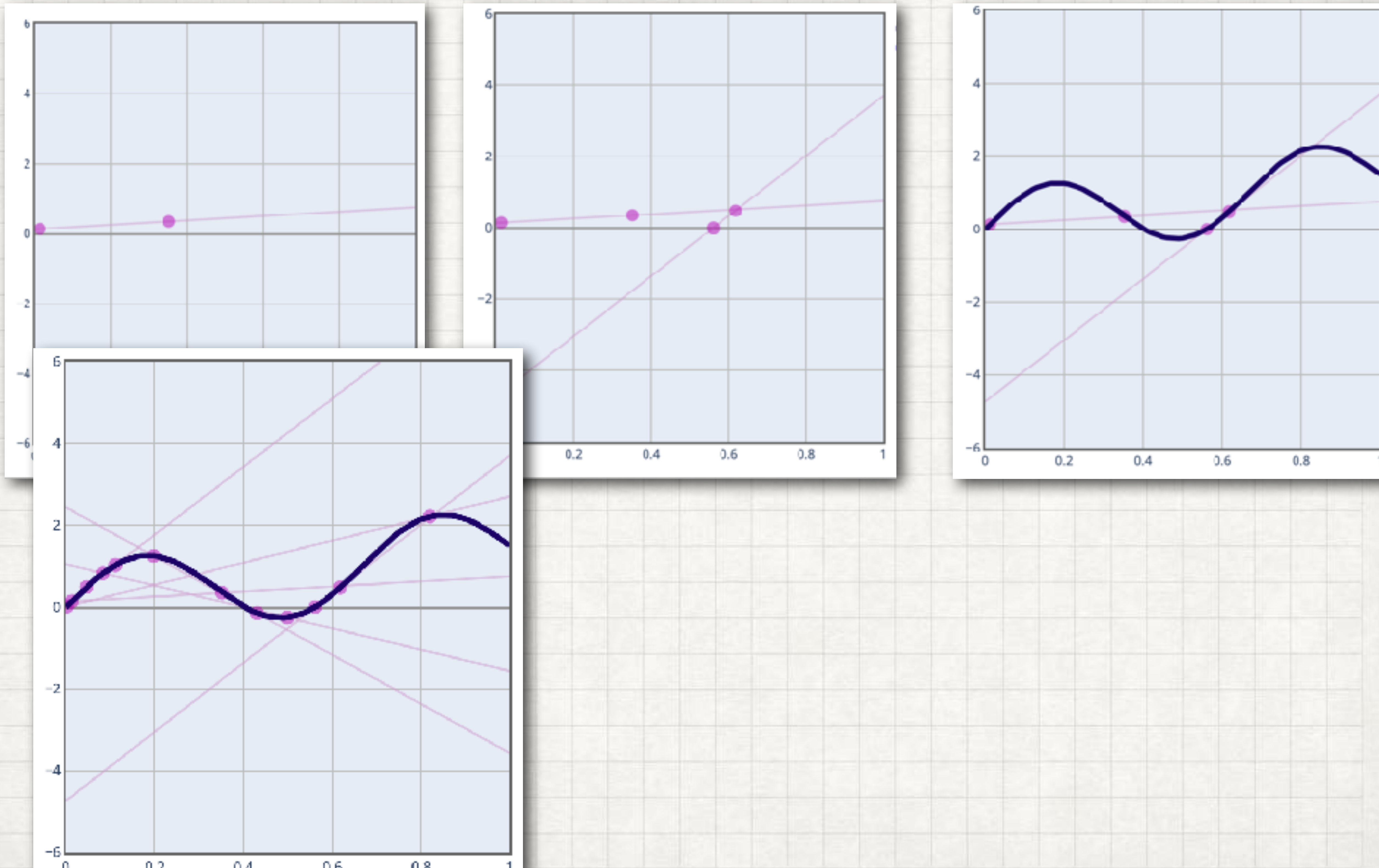
UNCERTAINTY IN LEARNING



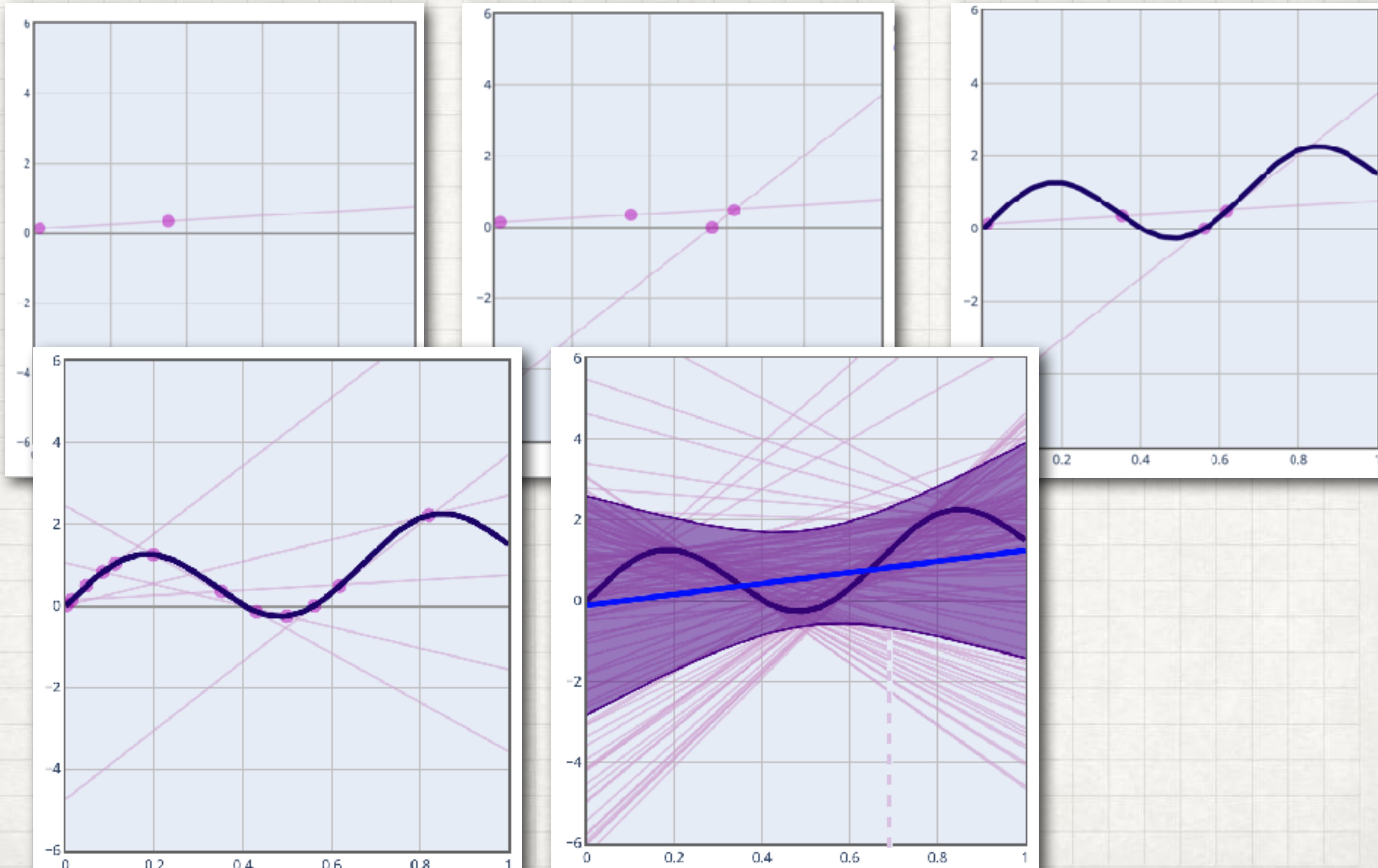




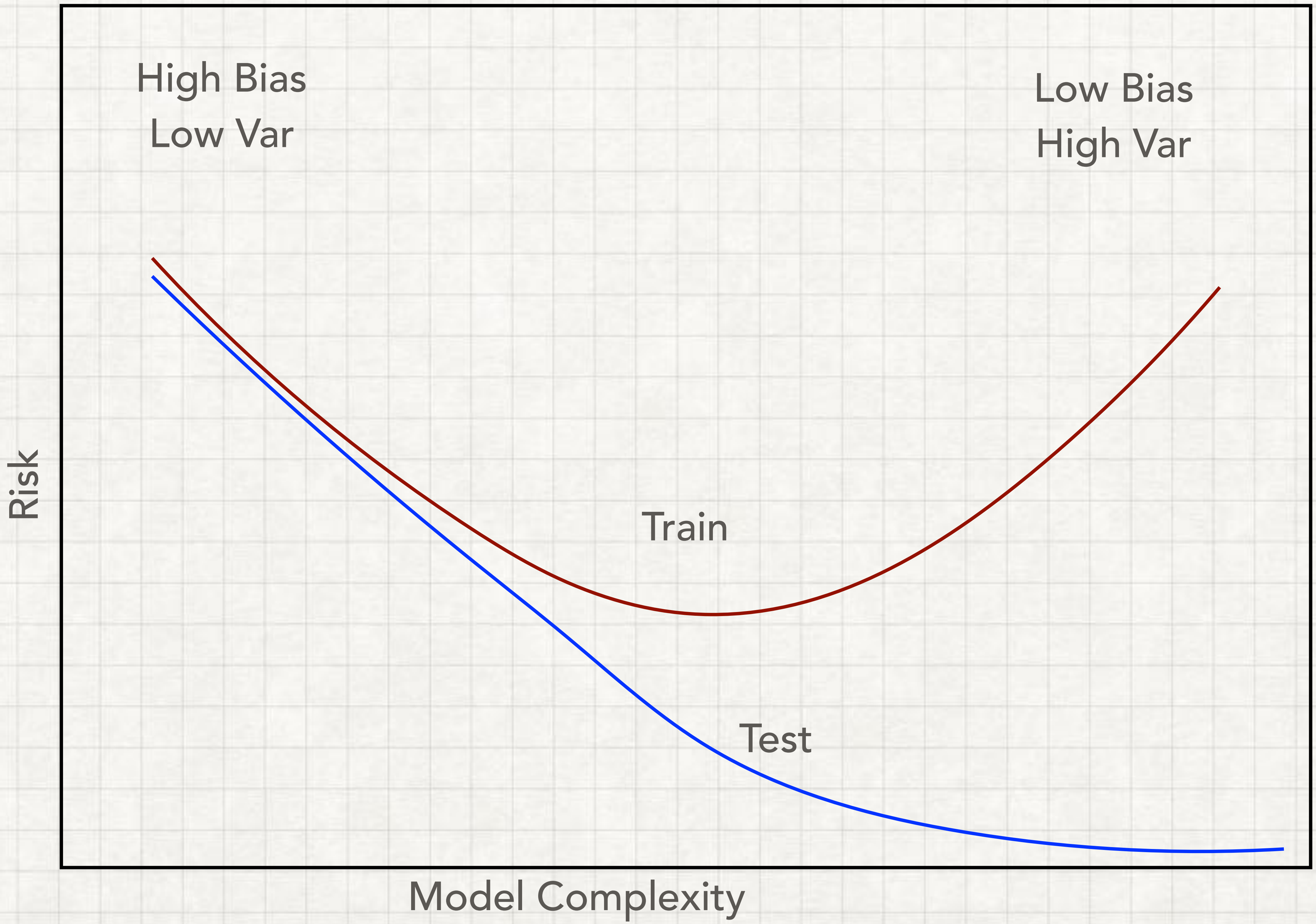
UNCERTAINTY IN LEARNING



UNCERTAINTY IN LEARNING



BIAS VS VARIANCE

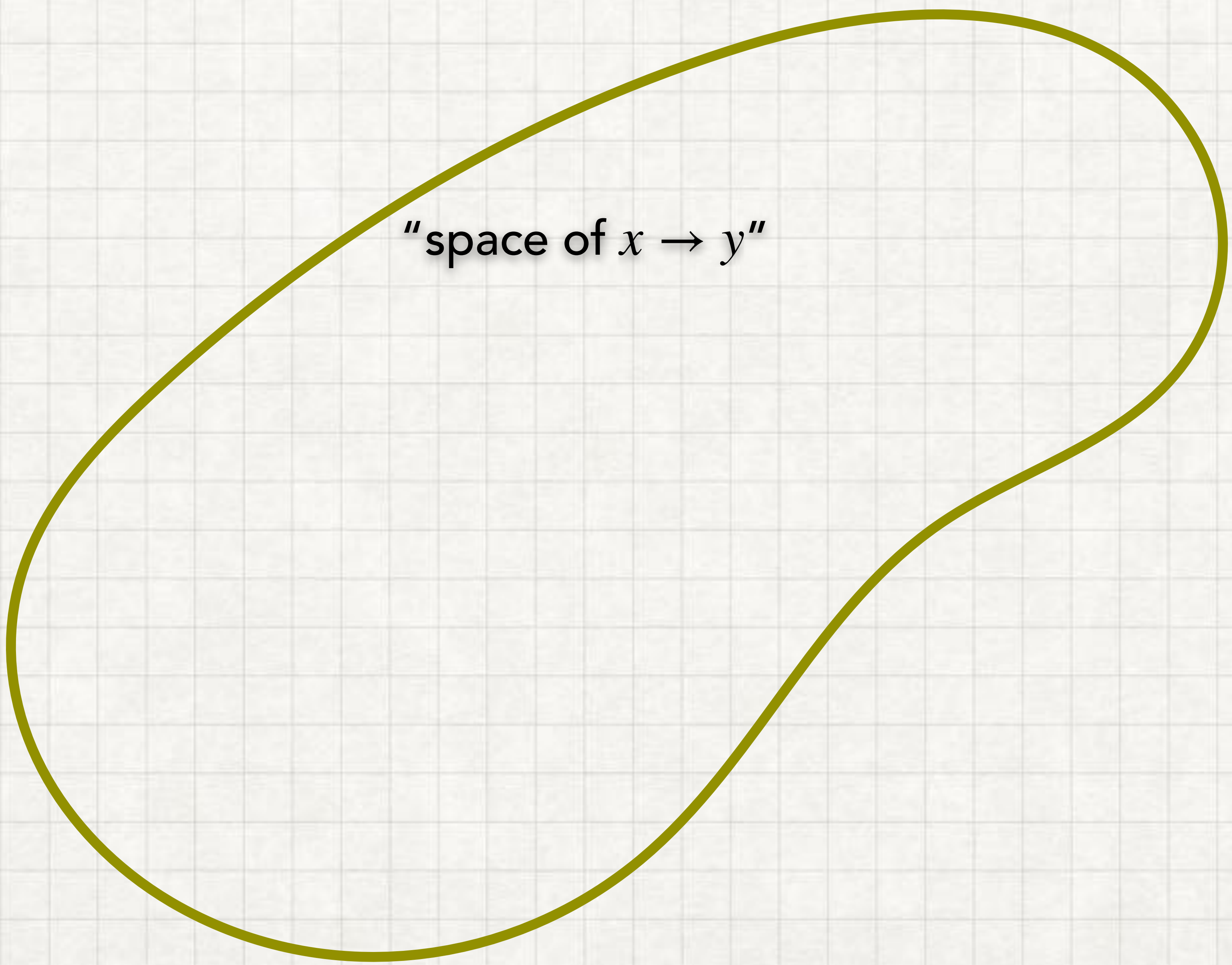


MOD1: INTUITIVE INTERPRETATION OF $\mathcal{H}$ FAMILY AND RISK

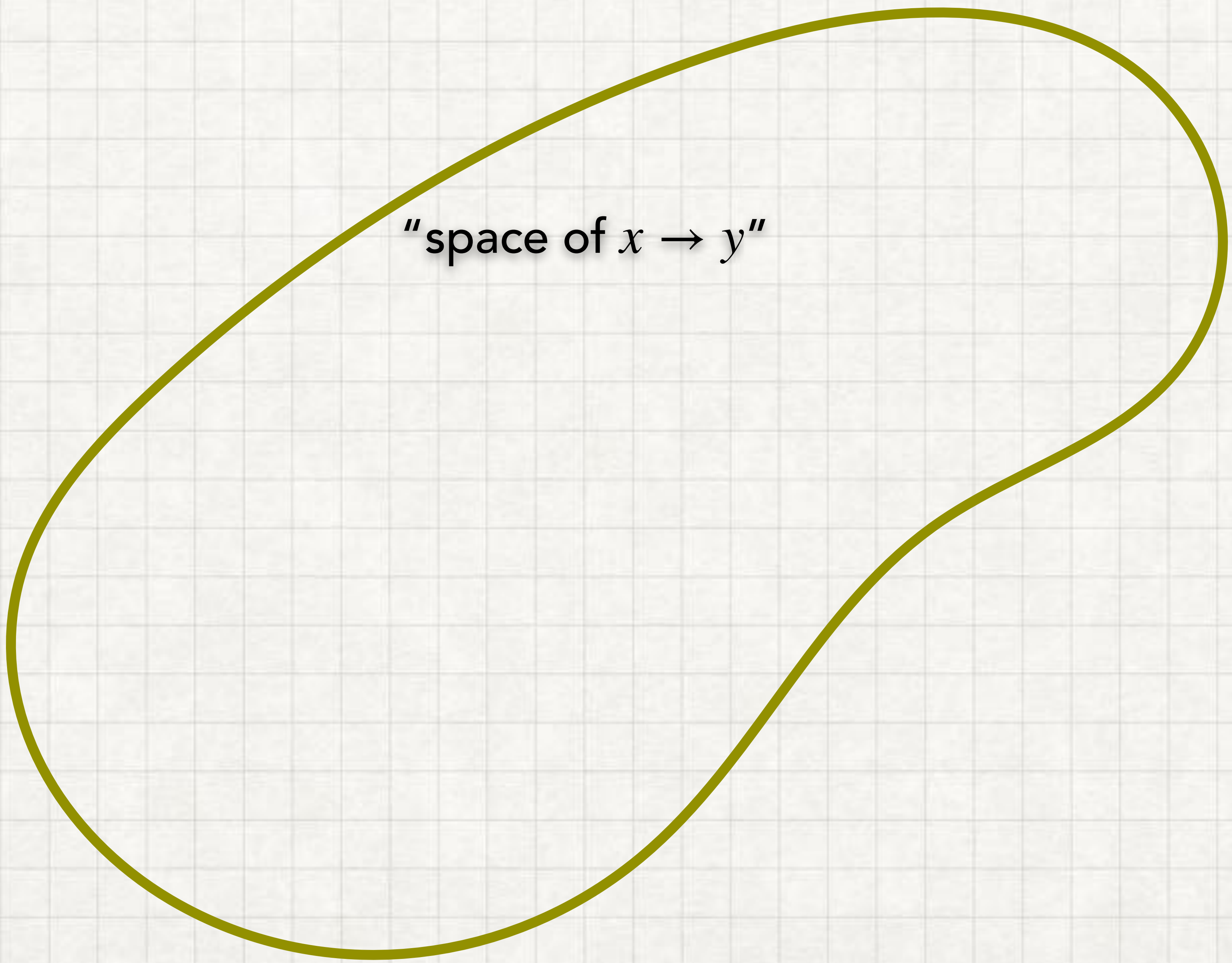
ERROR/RISK IN FUNCTIONAL SPACE

"space of $x \to y$"

ERROR/RISK IN FUNCTIONAL SPACE



FUNCTIONAL SPACE?

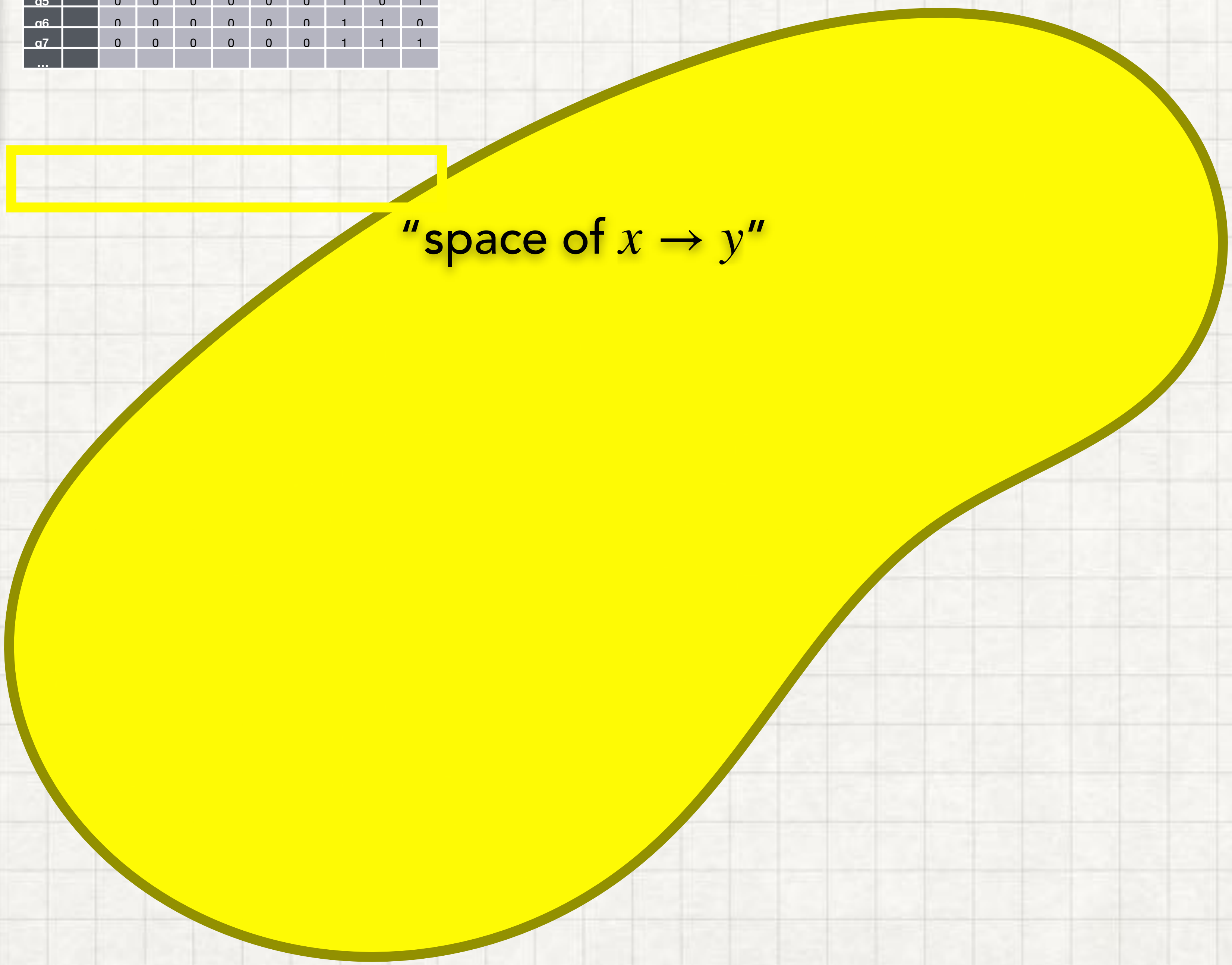


FUNCTIONAL SPACE?

- $2^{2500}=$

375828023454801203683362418972386504867736551759258677056523839782231681498337708535732725752658844333702457749526057760309227891351617765651907310968780236464694043316236562146724416478591131832593729111221580180531749232777515579969899075142213969117994877343802049421624954402214529390781647563339535024772584901607666862982567918622849636160208877365834950163790188523026247440507390382032188892386109905869706753143243921198482212075444022433366554786856559389689585638126582377224037721702239991441466026185752651502936472280911018500320375496336749951569521541850441747925844066295279671872605285792552660130702047998218334749356321677469529682551765858267502715894007887727250070780350262952377214028842297486263597879792176338220932619489509376

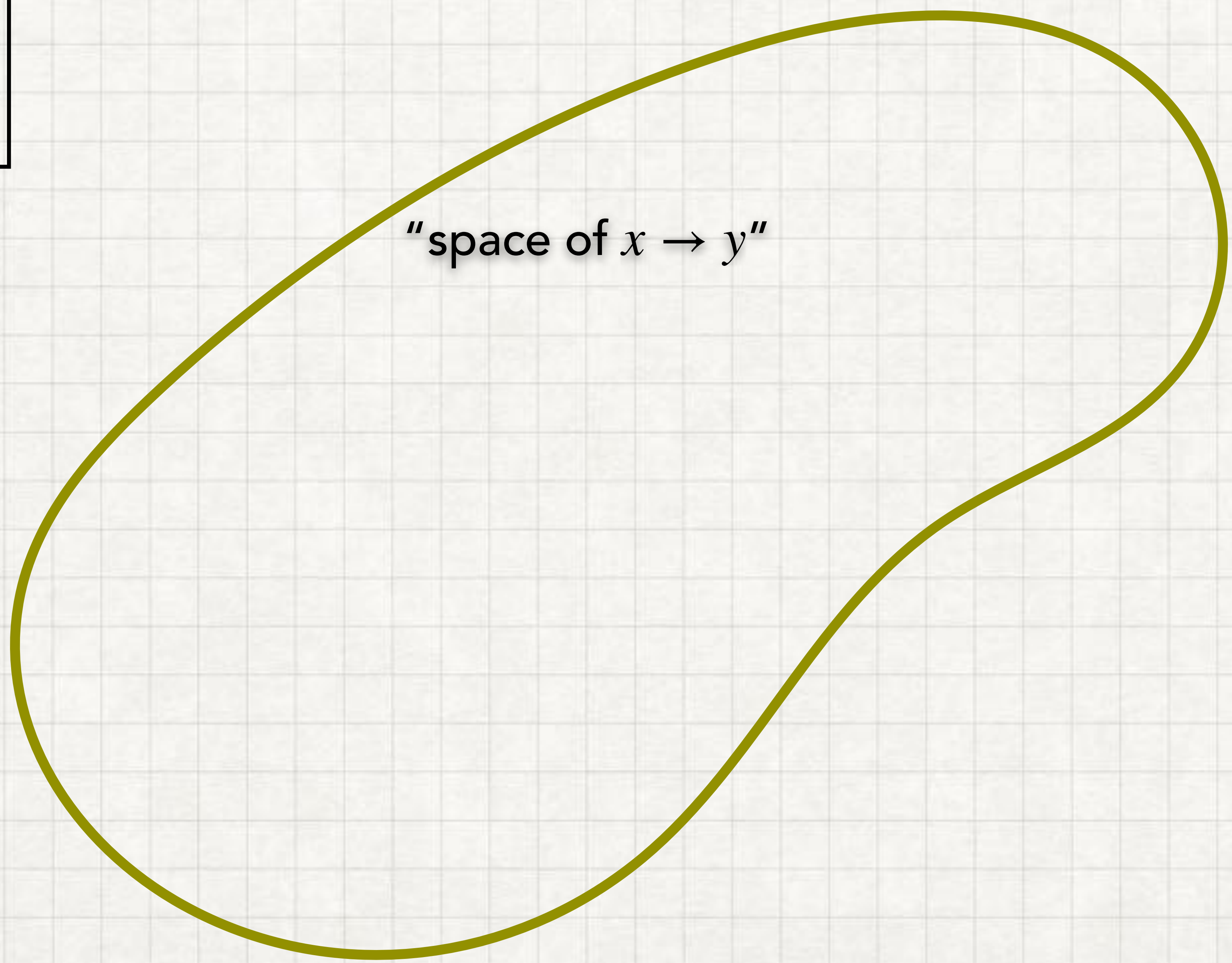
	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
g0		0	0	0	0	0	0	0	0	0
g1		0	0	0	0	0	0	0	0	1
g2		0	0	0	0	0	0	0	1	0
g3		0	0	0	0	0	0	0	1	1
g4		0	0	0	0	0	0	1	0	0
g5		0	0	0	0	0	0	1	0	1
g6		0	0	0	0	0	0	1	1	0
g7		0	0	0	0	0	0	1	1	1
...										



FUNCTIONAL SPACE ...

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

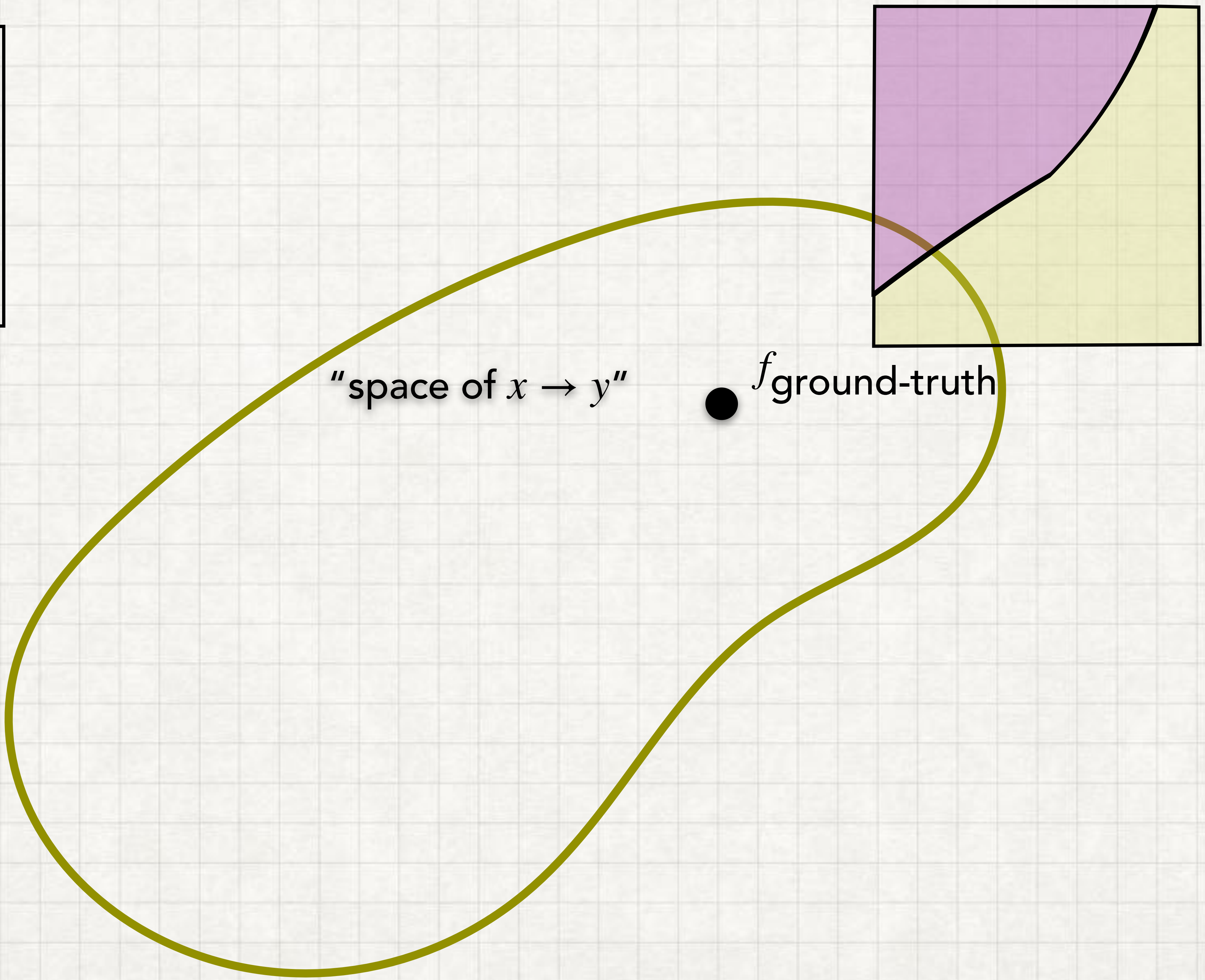
setosa / versicolor ?



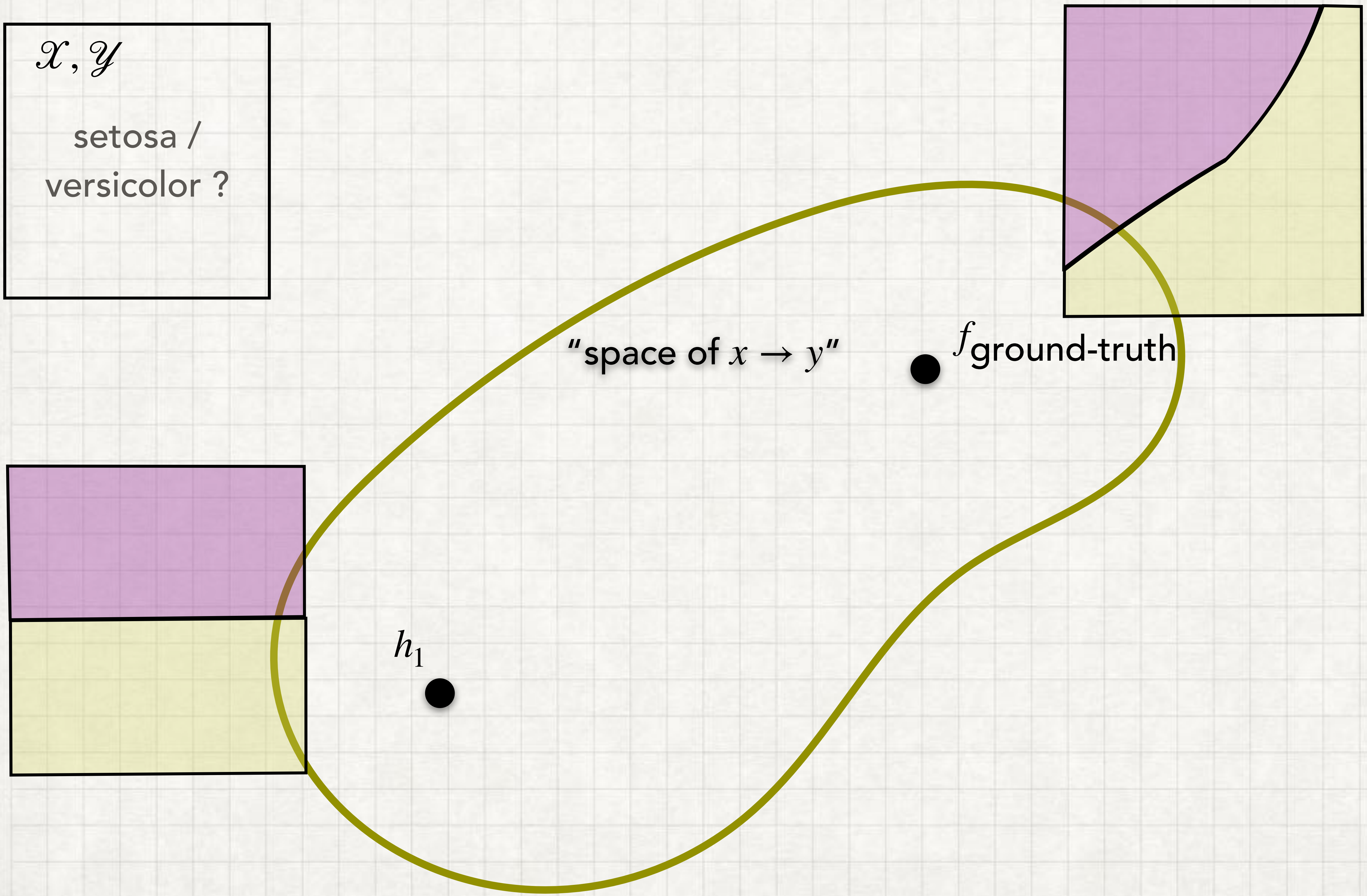
FUNCTIONAL SPACE …

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

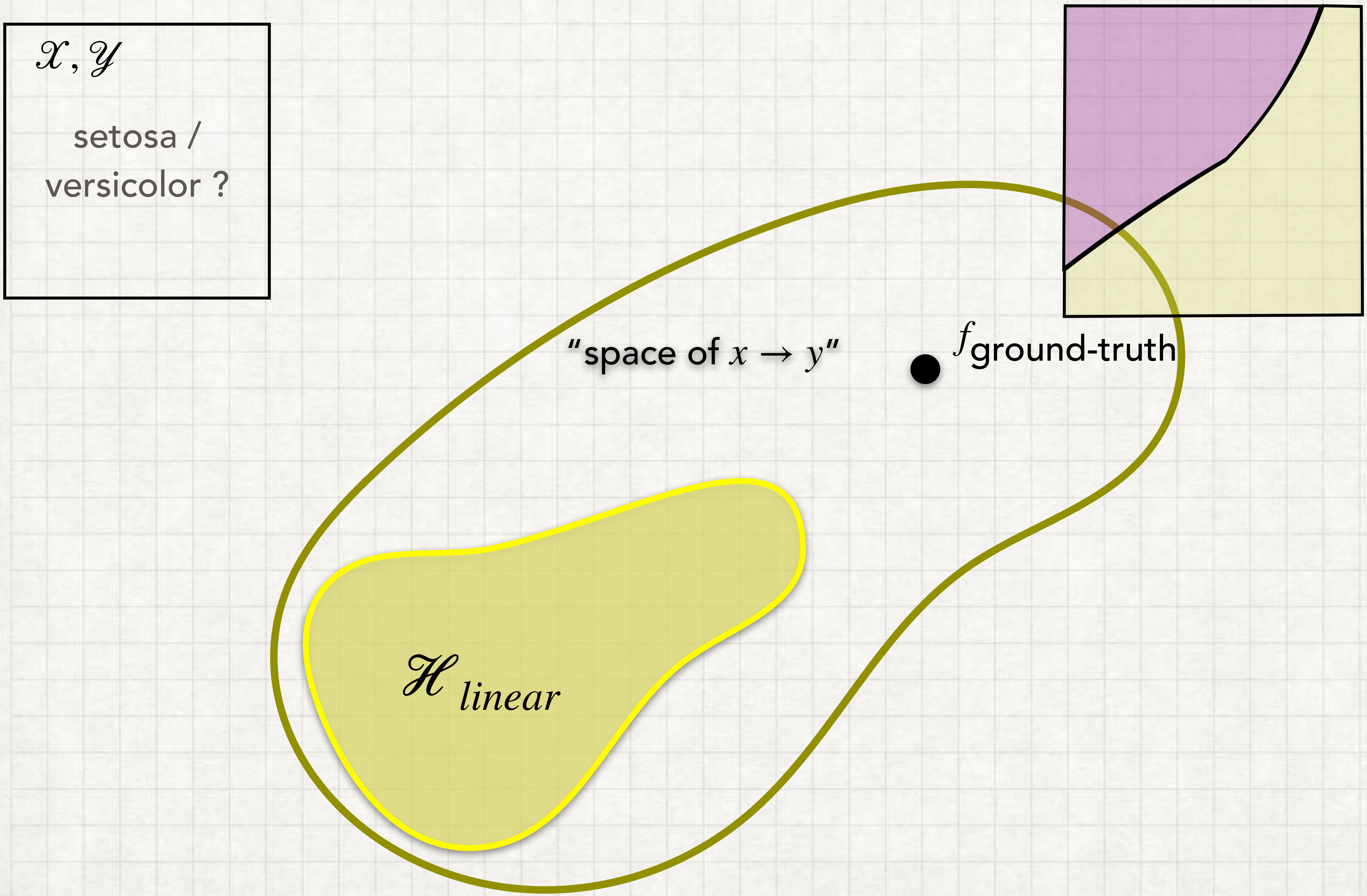
setosa / versicolor ?



FUNCTIONAL SPACE …



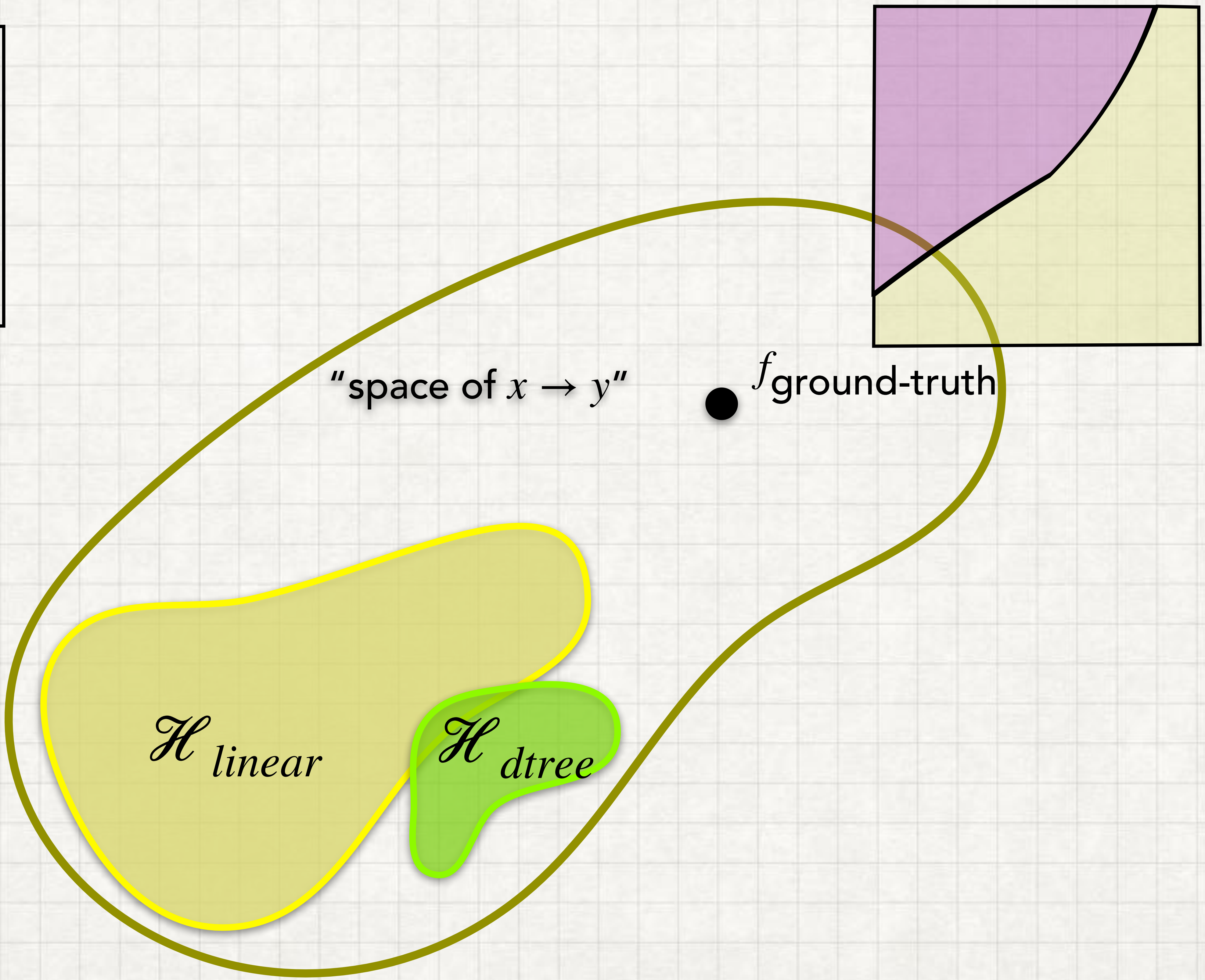
FUNCTIONAL SPACE.



FUNCTIONAL SPACE.

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

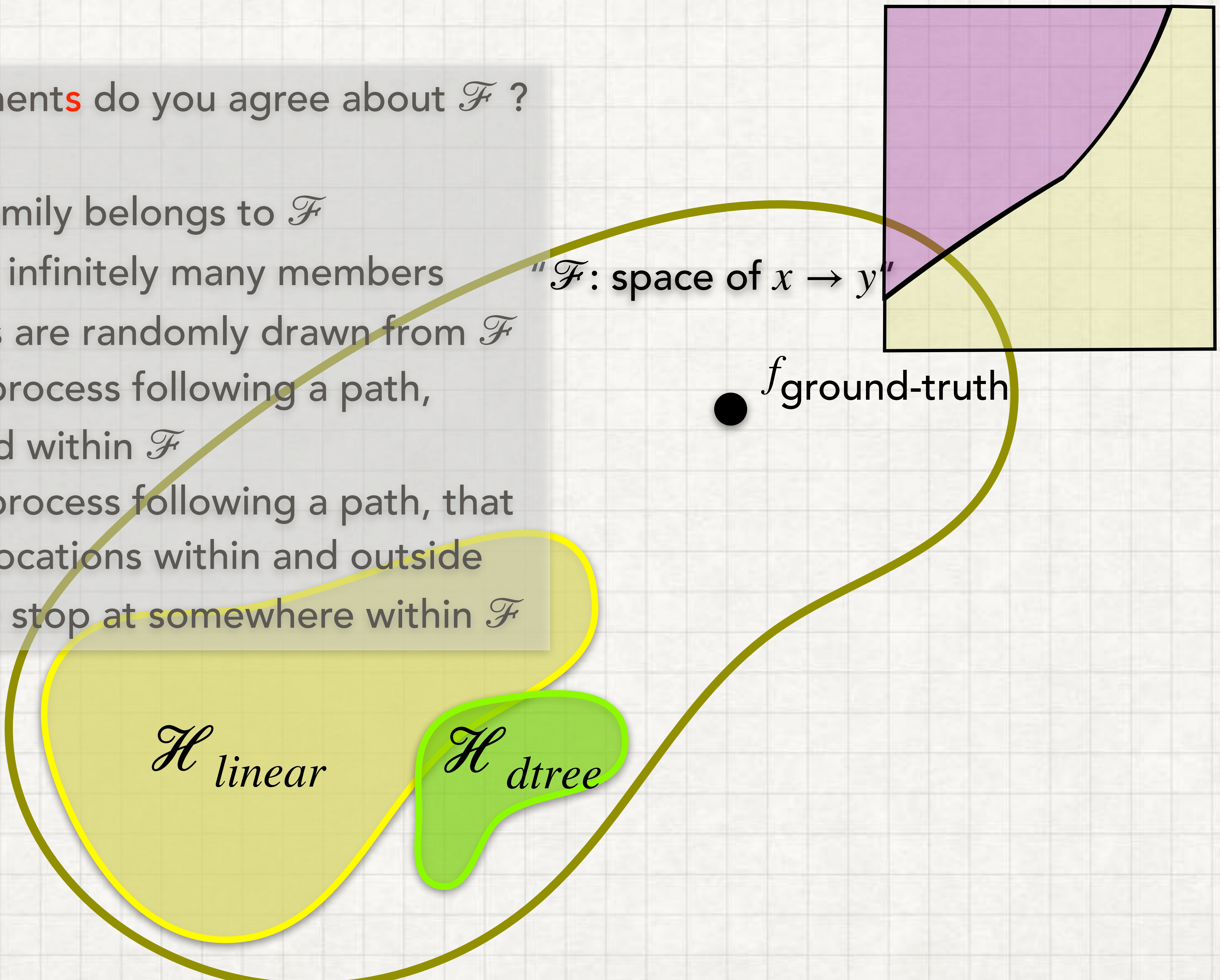
setosa /
versicolor ?



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

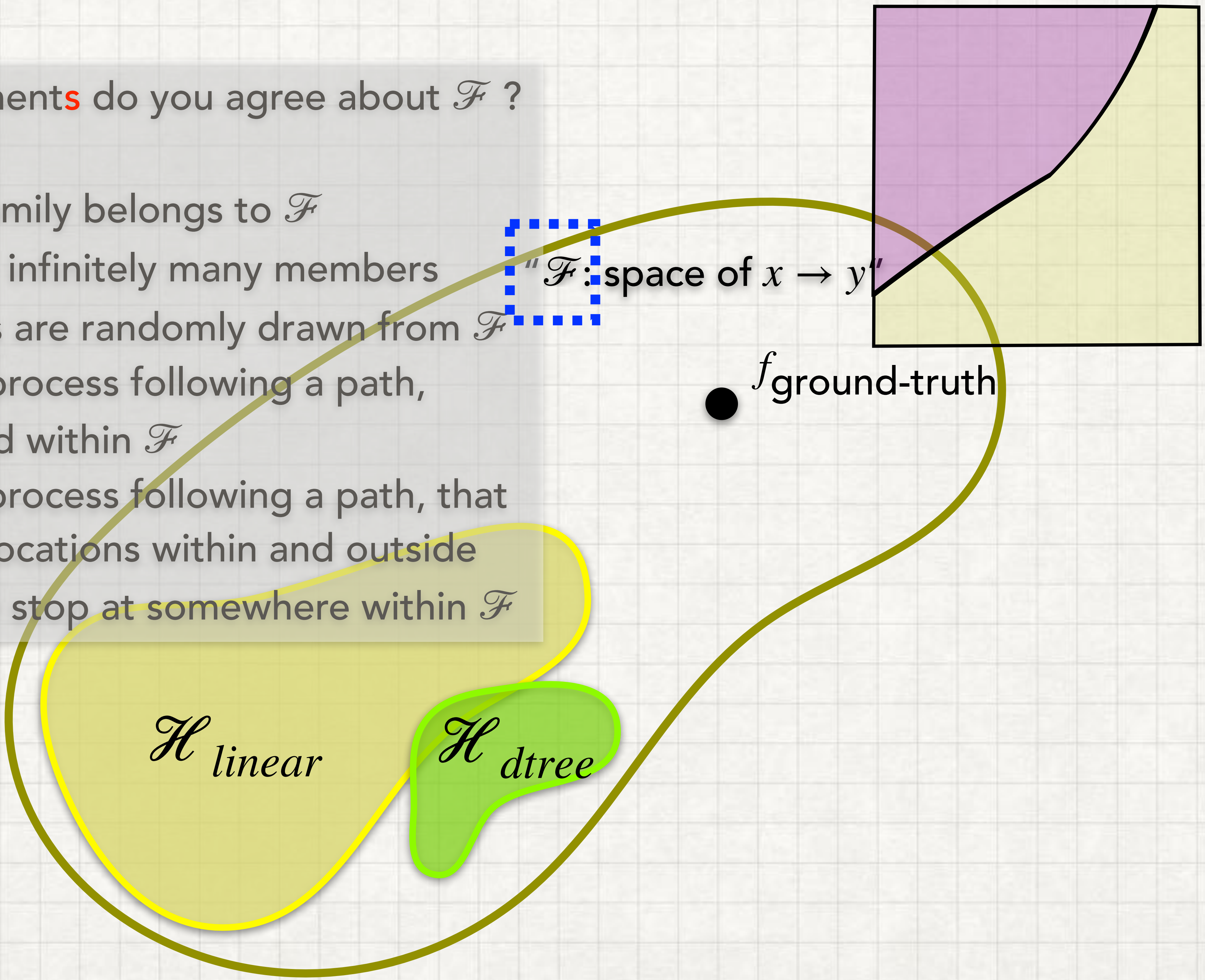
A. Any model family belongs to $\mathscr{F}$
B. $\mathscr{F}$ must have infinitely many members
C. Data samples are randomly drawn from $\mathscr{F}$
D. Training is a process following a path, strictly limited within $\mathscr{F}$
E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$
B. $\mathscr{F}$ must have infinitely many members
C. Data samples are randomly drawn from $\mathscr{F}$
D. Training is a process following a path, strictly limited within $\mathscr{F}$
E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$

B. $\mathscr{F}$ must have infinitely many members

C. Data samples are randomly drawn from $\mathscr{F}$

D. Training is a process following a path, strictly limited within $\mathscr{F}$

E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$

"$\mathscr{F}$: space of $x \rightarrow y$"

FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$

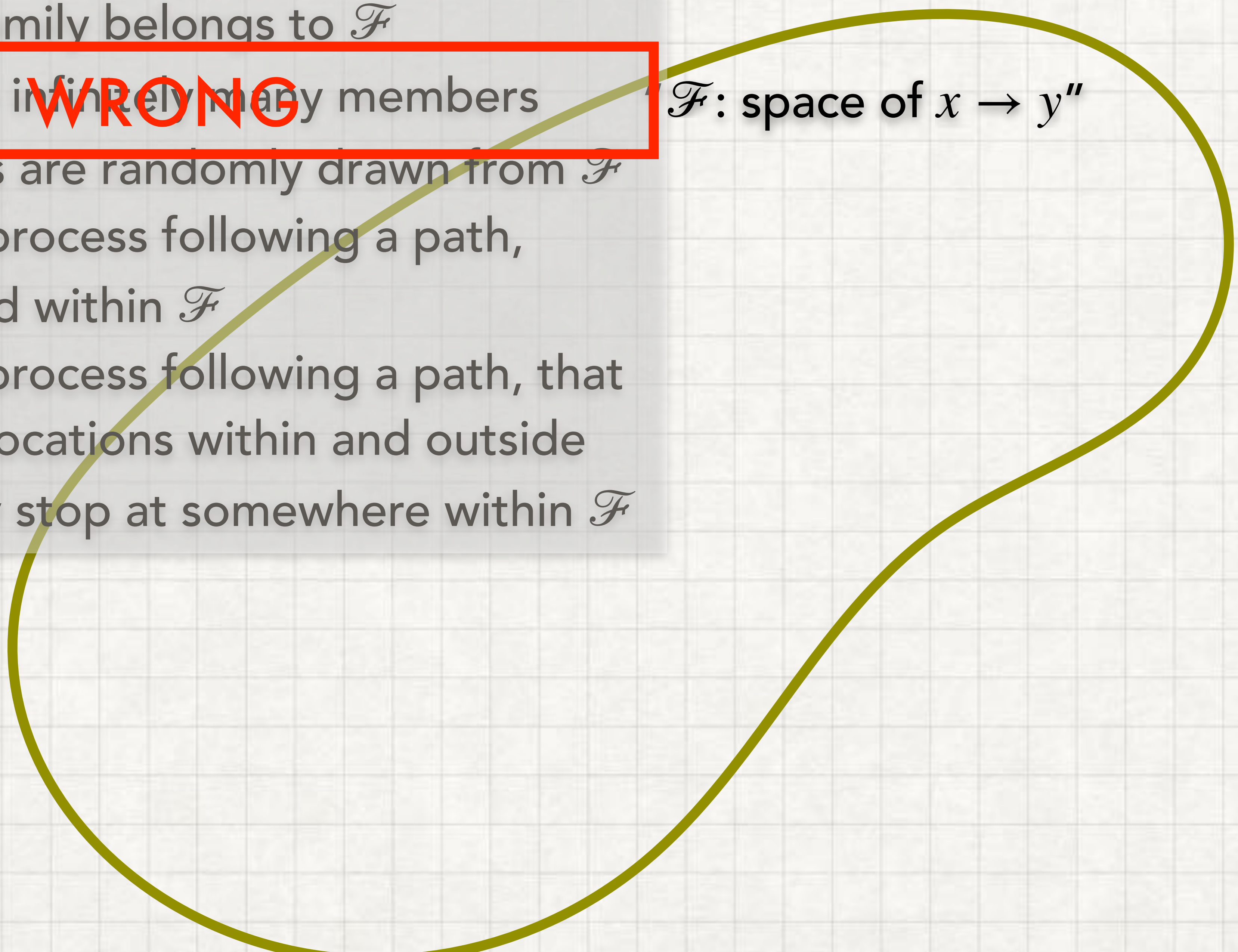
B. $\mathscr{F}$ must have infinitely many members

WRONG

C. Data samples are randomly drawn from $\mathscr{F}$

D. Training is a process following a path, strictly limited within $\mathscr{F}$

E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$
B. $\mathscr{F}$ must have infinitely many members

WRONG

C. Data samples are randomly drawn from $\mathscr{F}$
D. Training is a process following a path, strictly limited within $\mathscr{F}$
E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$

"$\mathscr{F}$: space of $x \to y$"

- $2^{2500}=$

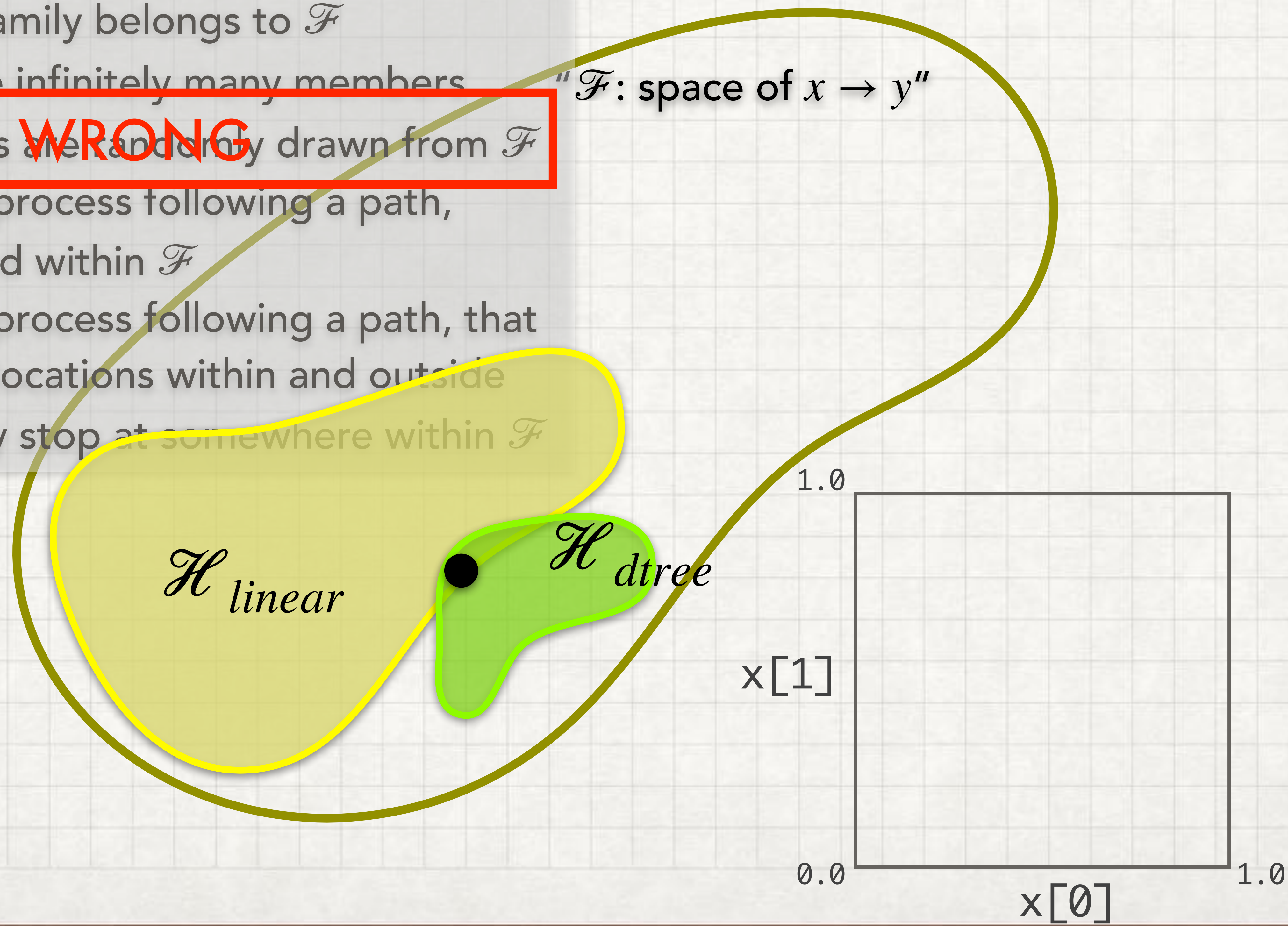
3758280234548012036833624189723865048677365517592586770565238397822316814983377085357327257526588443337024577495260577603092278913516177656519073109687802364646940433162365621467244164785911318325937291112215801805317492327775155799698990751422139691179948773438020494216249544022145293907816475633395350247725849016076668629825679186228496361602088773658349501637901885230262474405073903820321888923861099058697067531432439211984822120754440224333665547868565593896895856381265823772240377217022399914414660261857526515029364722809110185003203754963367499515695215418504417479258440662952796718726052857925526601307020479982183347493563216774695296825517658582675027158940078877272500707803502629523772140288422974862635978797921763382209326194895093 76

	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
g0		0	0	0	0	0	0	0	0	0
g1		0	0	0	0	0	0	0	0	1
g2		0	0	0	0	0	0	0	1	0
g3		0	0	0	0	0	0	0	1	1
g4		0	0	0	0	0	0	1	0	0
g5		0	0	0	0	0	0	1	0	1
g6		0	0	0	0	0	0	1	1	0
g7		0	0	0	0	0	0	1	1	1
...										

FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$
B. $\mathscr{F}$ must have infinitely many members
C. Data samples are randomly drawn from $\mathscr{F}$ WRONG
D. Training is a process following a path, strictly limited within $\mathscr{F}$
E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

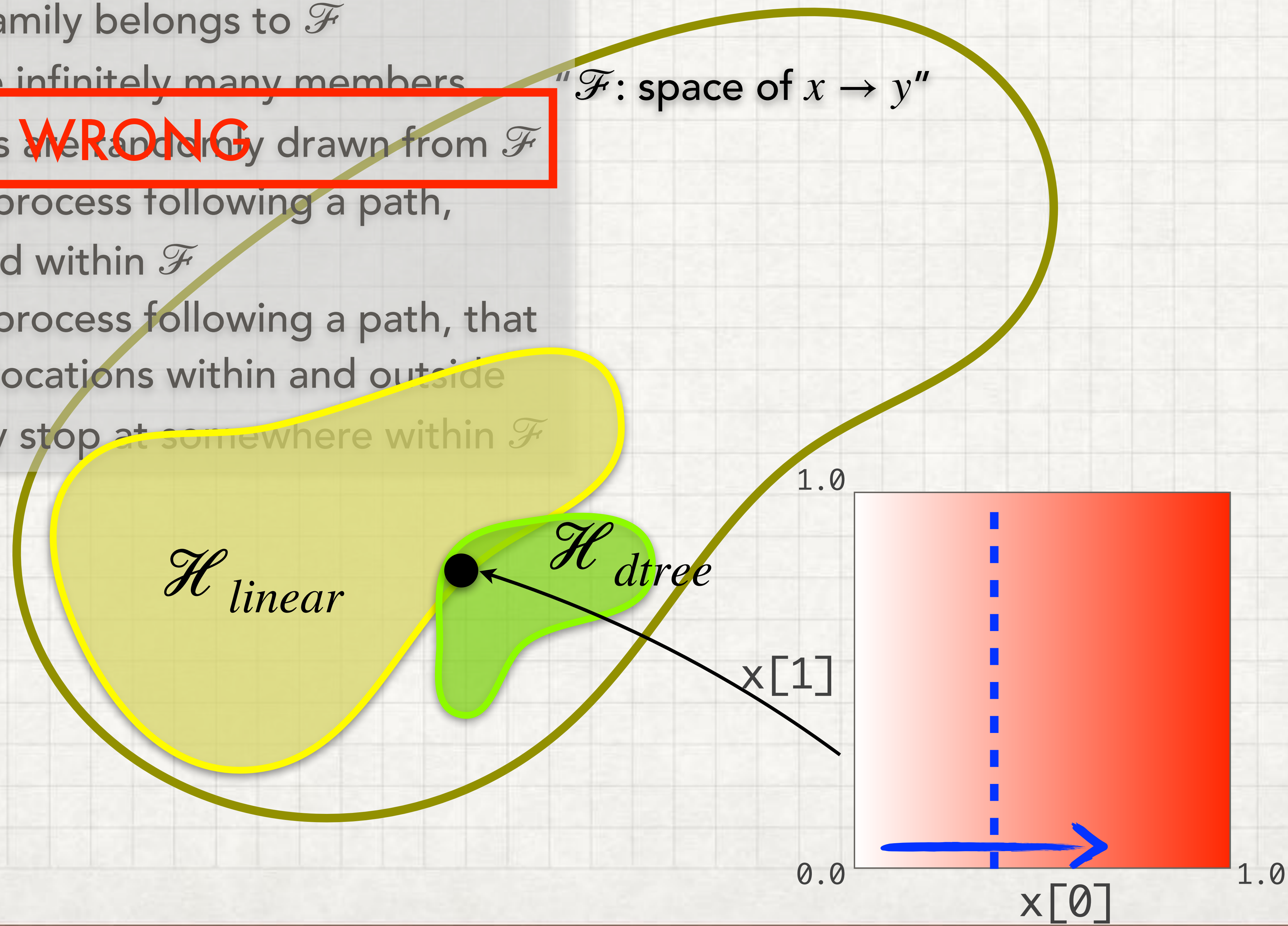
A. Any model family belongs to $\mathscr{F}$

B. $\mathscr{F}$ must have infinitely many members

C. Data samples are randomly drawn from $\mathscr{F}$ WRONG

D. Training is a process following a path, strictly limited within $\mathscr{F}$

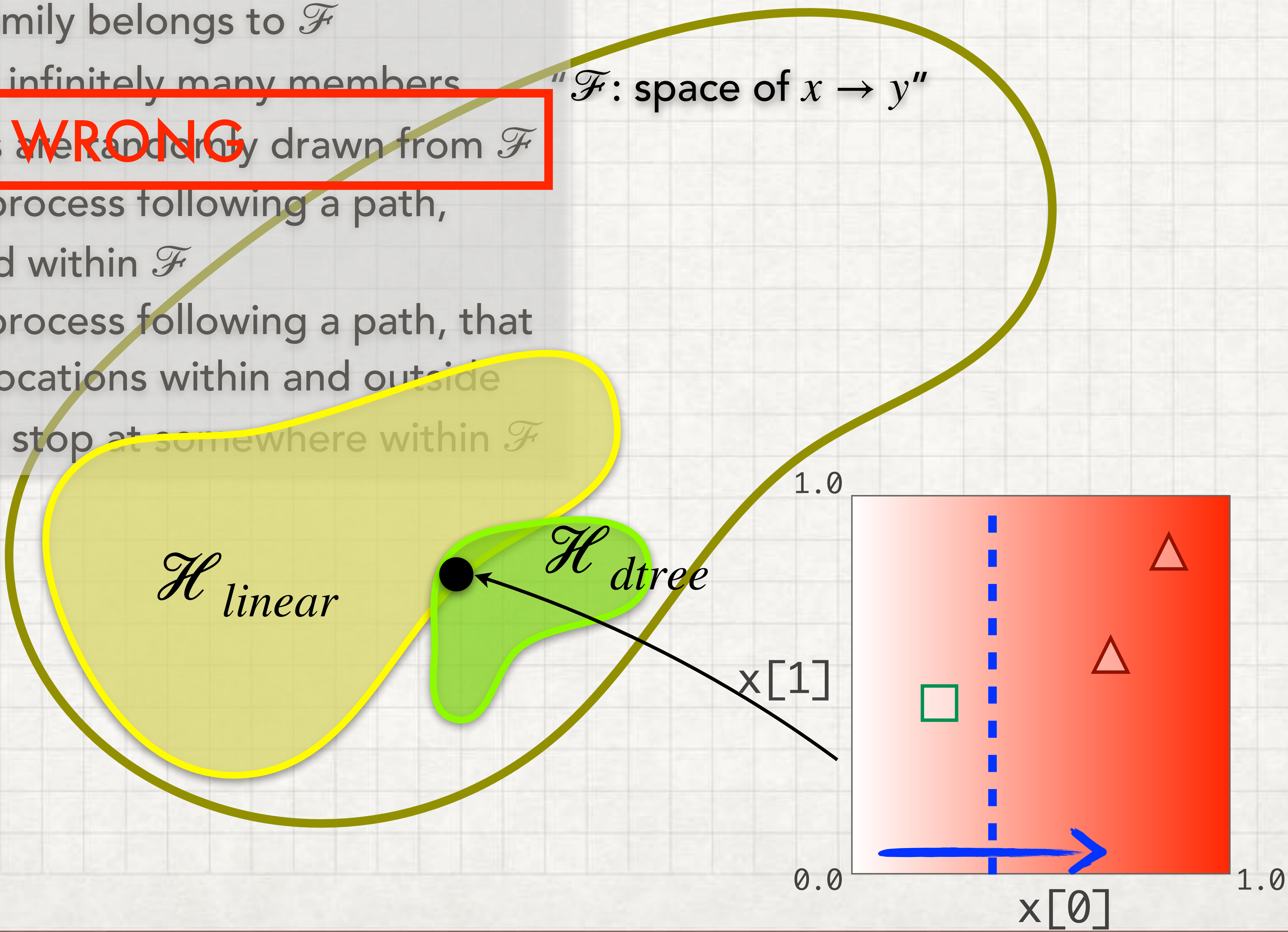
E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$
B. $\mathscr{F}$ must have infinitely many members
C. Data samples are randomly drawn from $\mathscr{F}$ WRONG
D. Training is a process following a path, strictly limited within $\mathscr{F}$
E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$



FUNCTIONAL SPACE.

Q: Which statements do you agree about $\mathscr{F}$?

A. Any model family belongs to $\mathscr{F}$

B. $\mathscr{F}$ must have infinitely many members

C. Data samples are randomly drawn from $\mathscr{F}$

D. Training is a process following a path, strictly limited within $\mathscr{F}$

E. Training is a process following a path, that may lead to locations within and outside $\mathscr{F}$, but finally stop at somewhere within $\mathscr{F}$

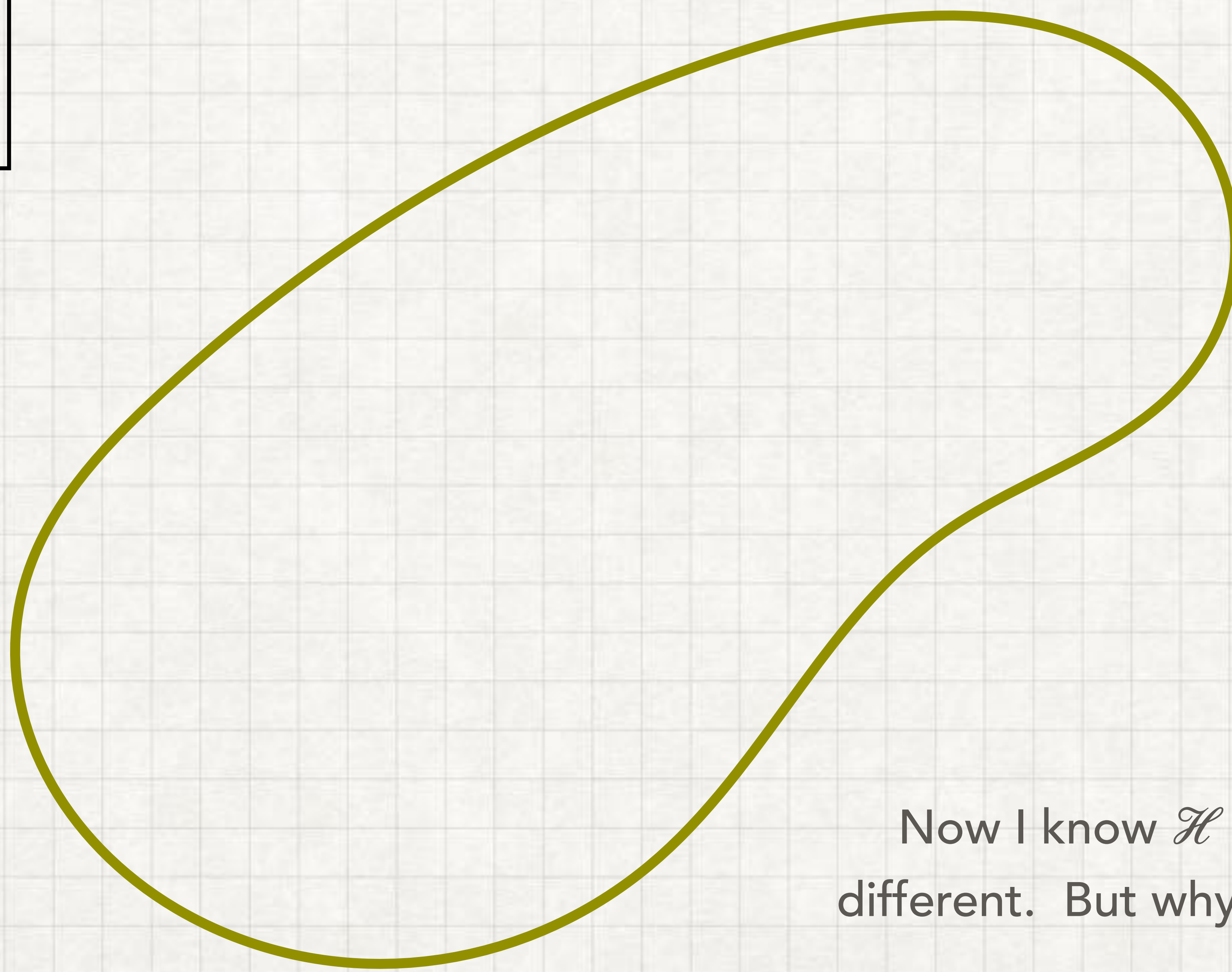
WRONG

"$\mathscr{F}$: space of $x \to y$"

WHY FUNCTIONAL SPACE?

$\mathscr{X}, \mathscr{Y}$

setosa /
versicolor ?

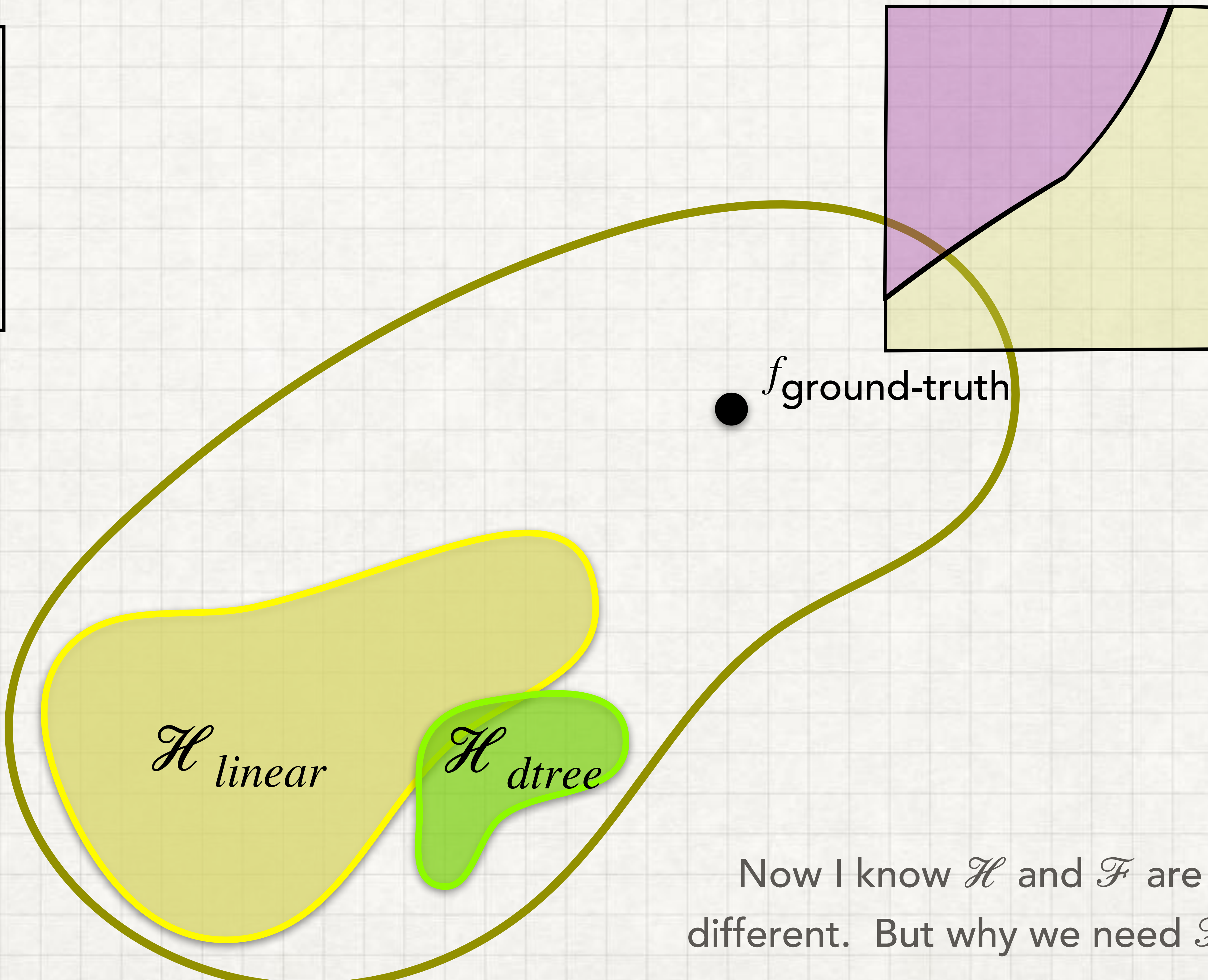


Now I know $\mathscr{H}$ and $\mathscr{F}$ are different. But why we need $\mathscr{F}$?

WHY FUNCTIONAL SPACE?

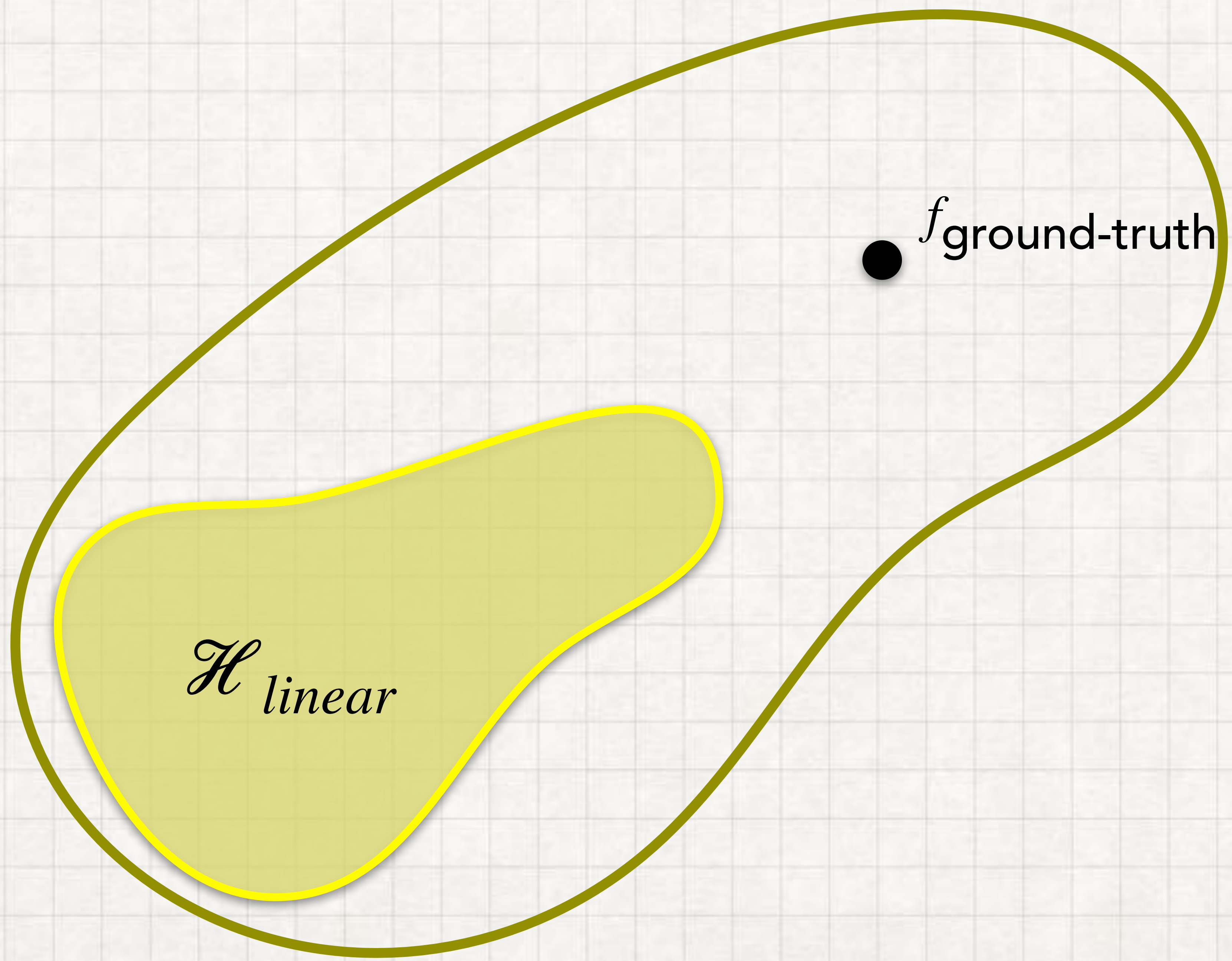
$\mathscr{X}, \mathscr{Y}$

setosa /
versicolor ?

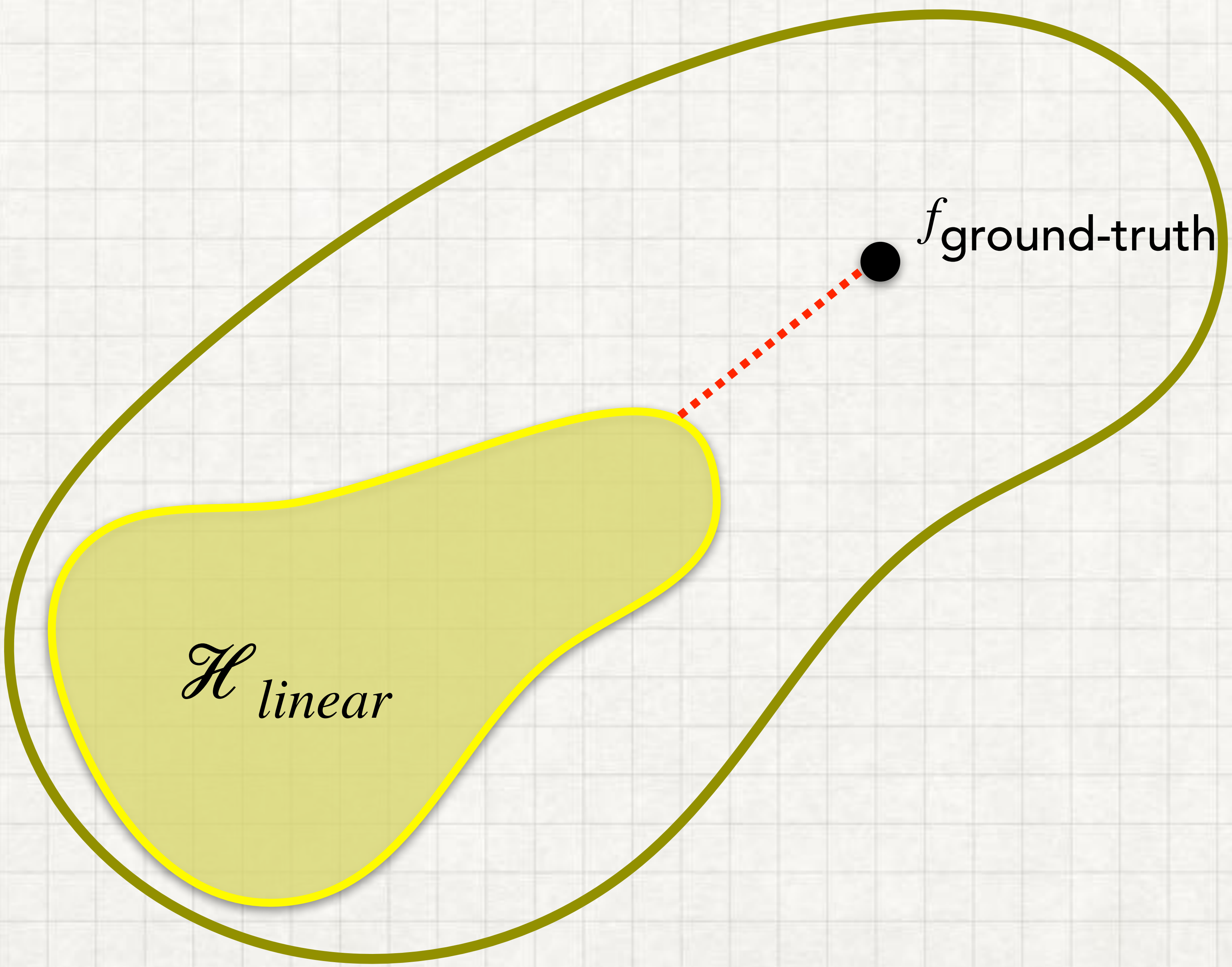


Now I know $\mathscr{H}$ and $\mathscr{F}$ are different. But why we need $\mathscr{F}$?

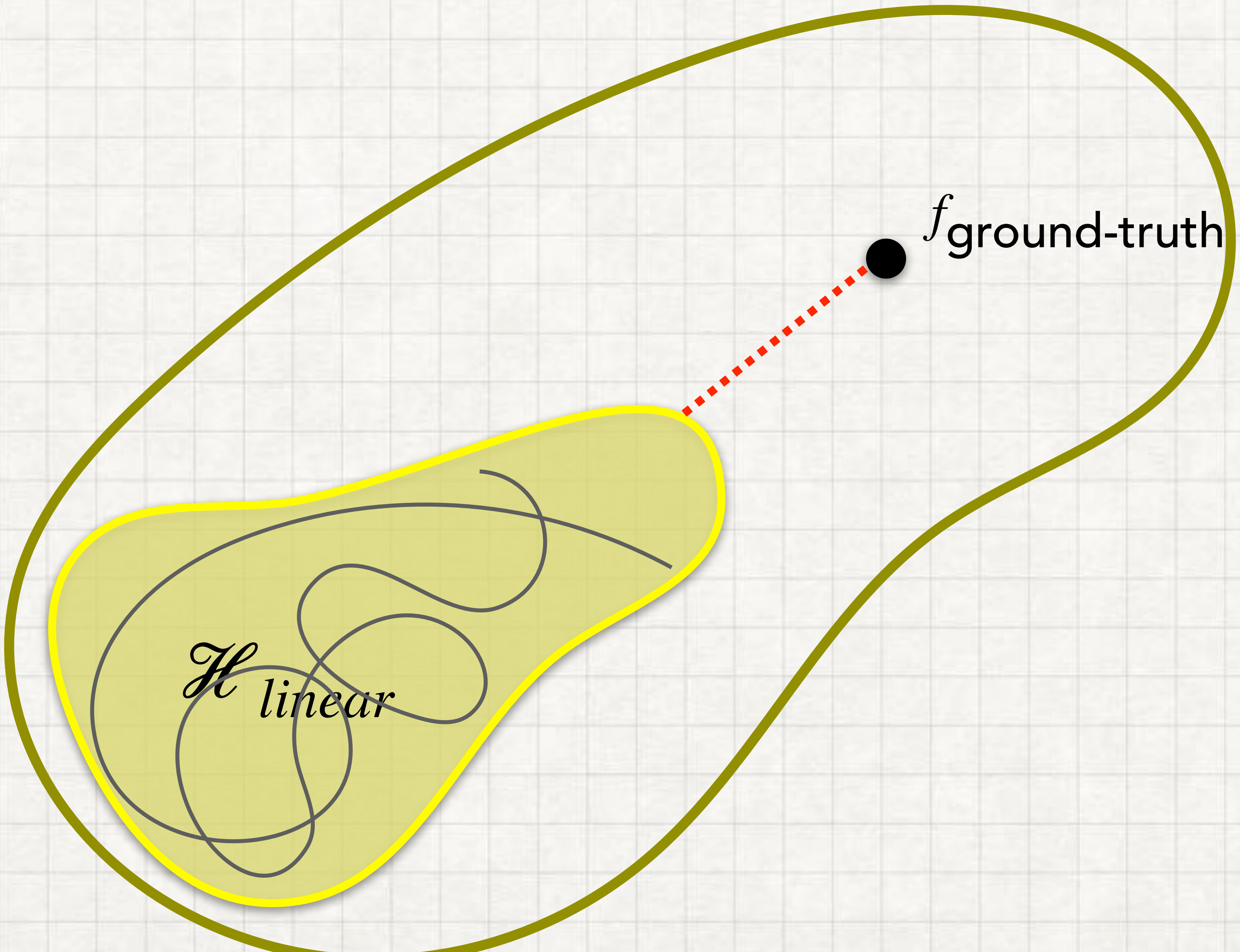
WHY FUNCTIONAL SPACE – GIVEN $\mathscr{H}$?



WHY FUNCTIONAL SPACE – GIVEN $\mathscr{H}$?



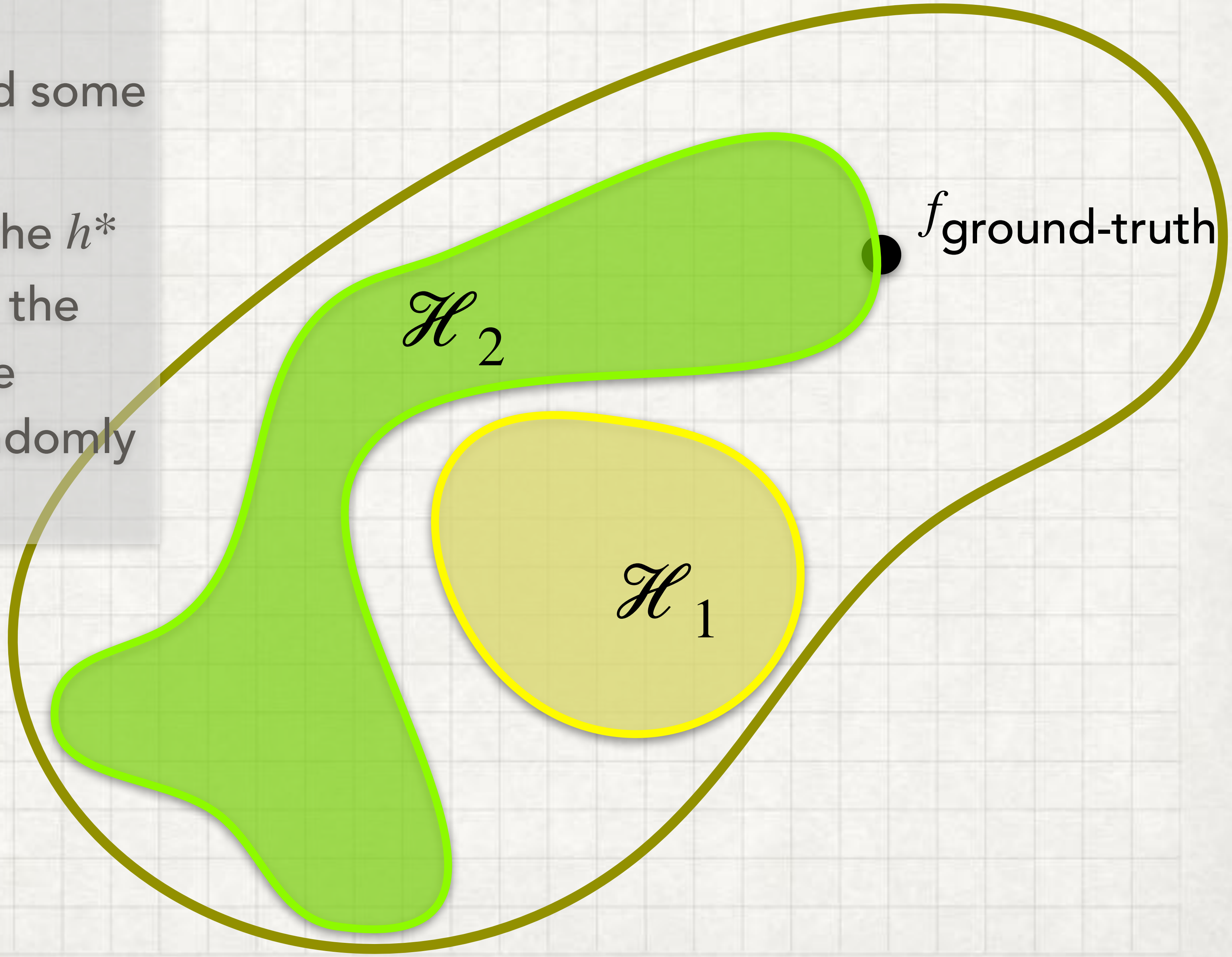
WHY FUNCTIONAL SPACE – GIVEN $\mathscr{H}$?



WHICH ONE IS BETTER?

Q: Which hypothesis family will result in less error?

A. $\mathscr{H}_2$, it (almost) includes the ground-truth, so one can find some good predictor h^* from $\mathscr{H}_2$
B. Not sure, because whether the h^* selected from $\mathscr{H}_2$ is close to the ground-truth depends on the training dataset, which is randomly generated.

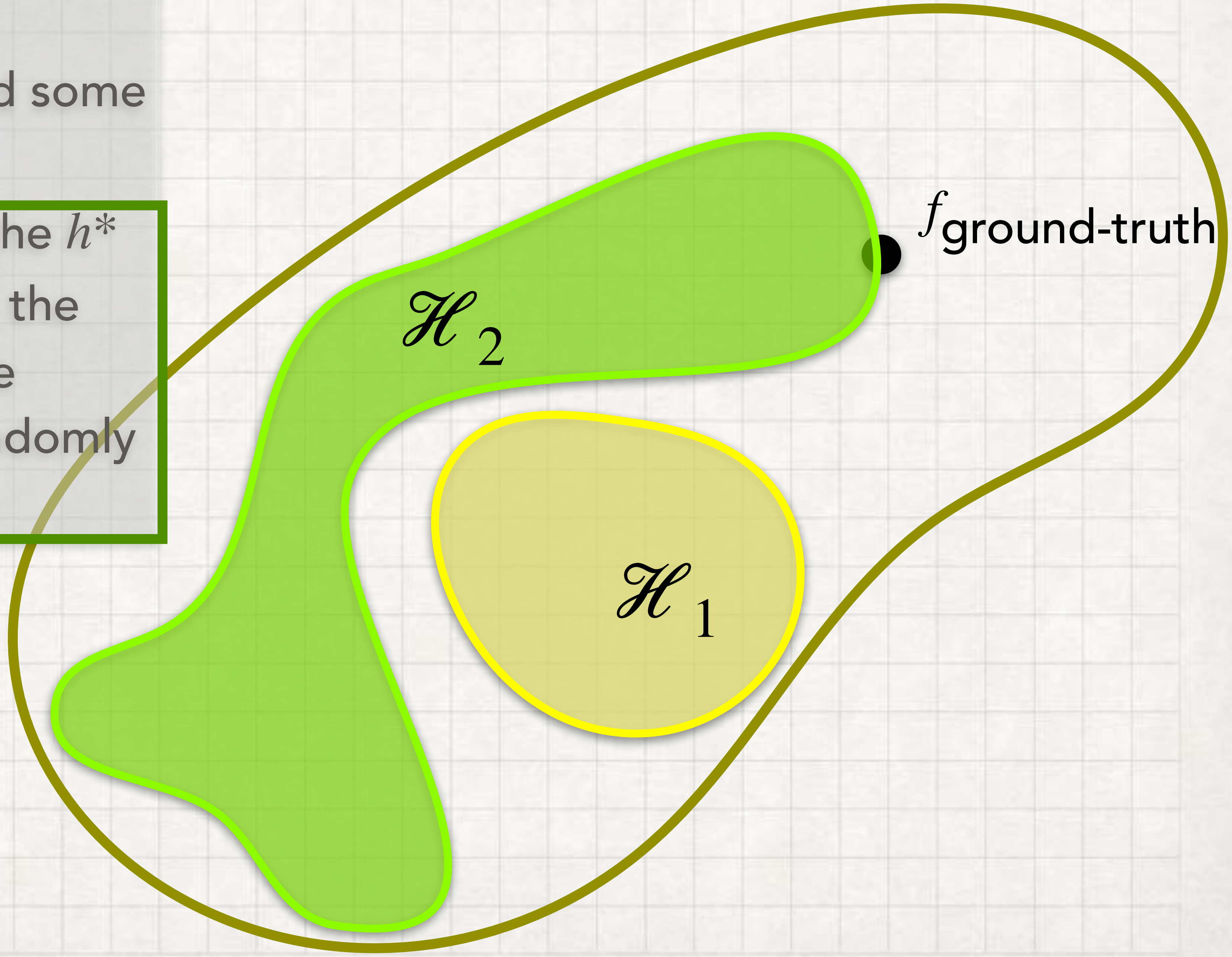


WHICH ONE IS BETTER?

Q: Which hypothesis family will result in less error?

A. $\mathscr{H}_2$, it (almost) includes the ground-truth, so one can find some good predictor h^* from $\mathscr{H}_2$

B. Not sure, because whether the h^* selected from $\mathscr{H}_2$ is close to the ground-truth depends on the training dataset, which is randomly generated.



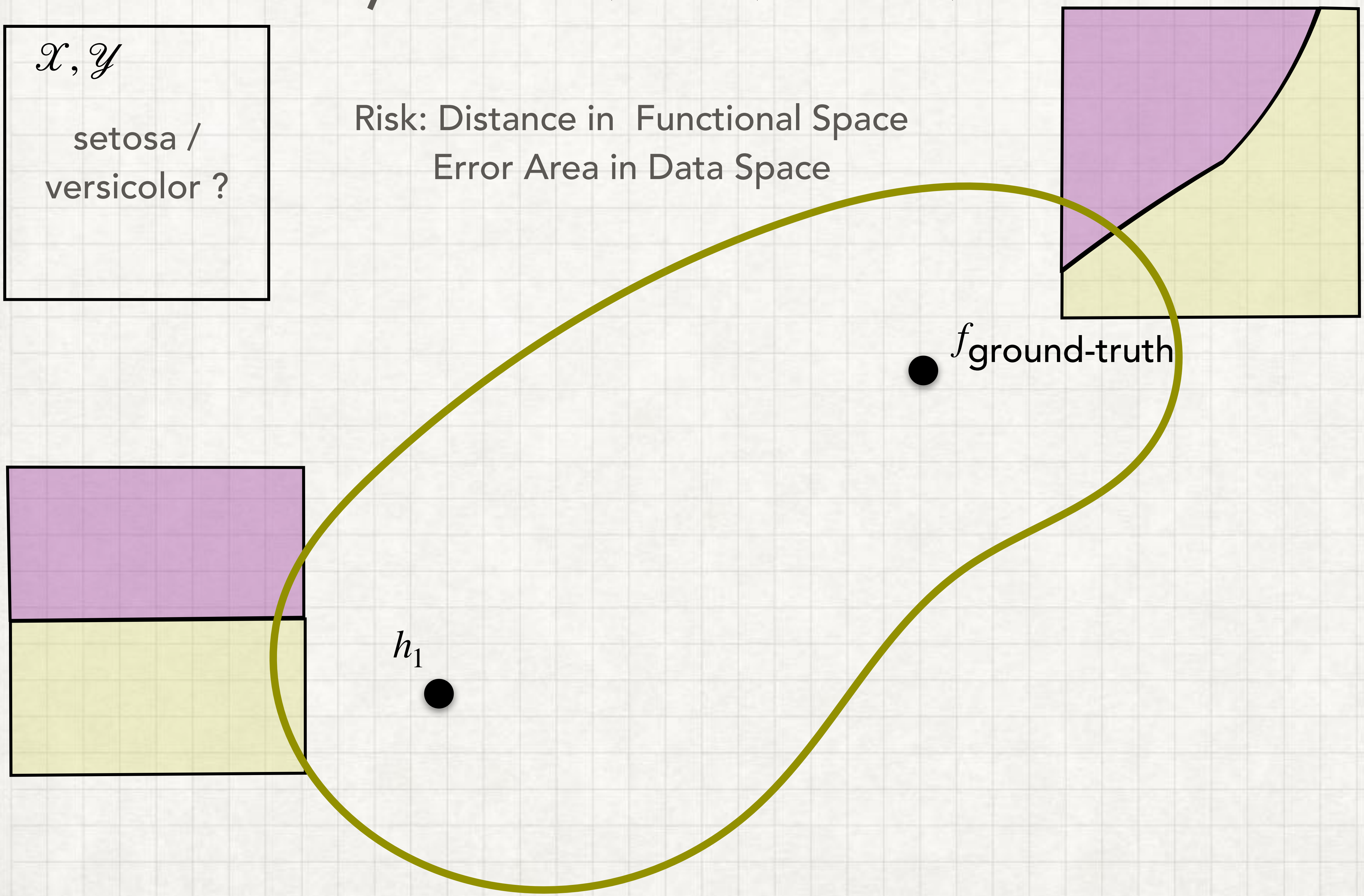
ERROR/RISK IN FUNCTIONAL SPACE

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

setosa / versicolor ?

Risk: Distance in Functional Space

Error Area in Data Space

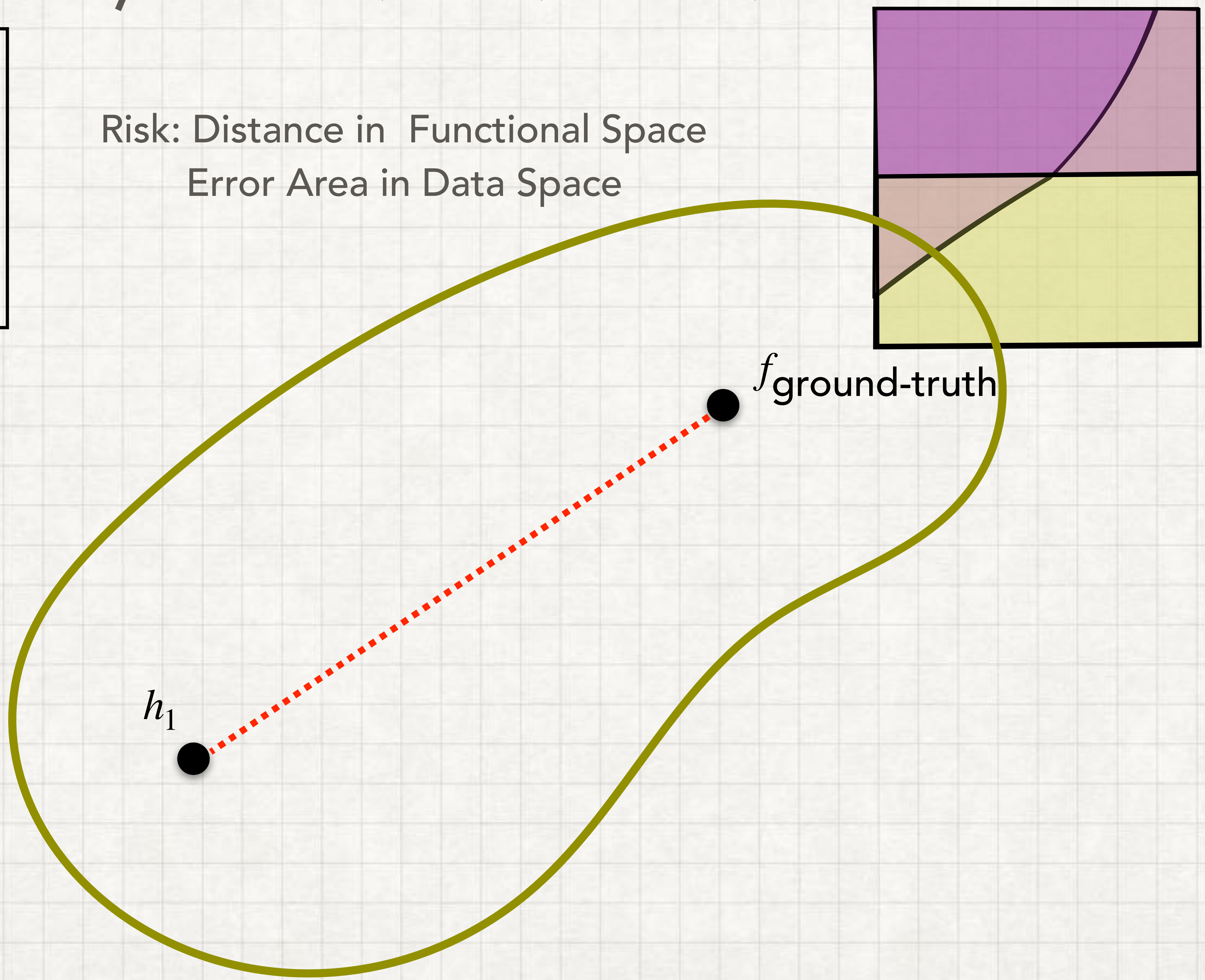


ERROR/RISK IN FUNCTIONAL SPACE

$\mathscr{X}, \mathscr{Y}$

setosa /
versicolor ?

Risk: Distance in Functional Space
Error Area in Data Space

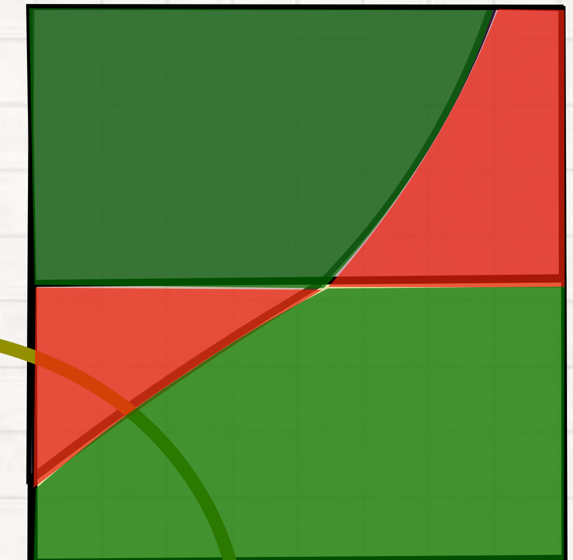


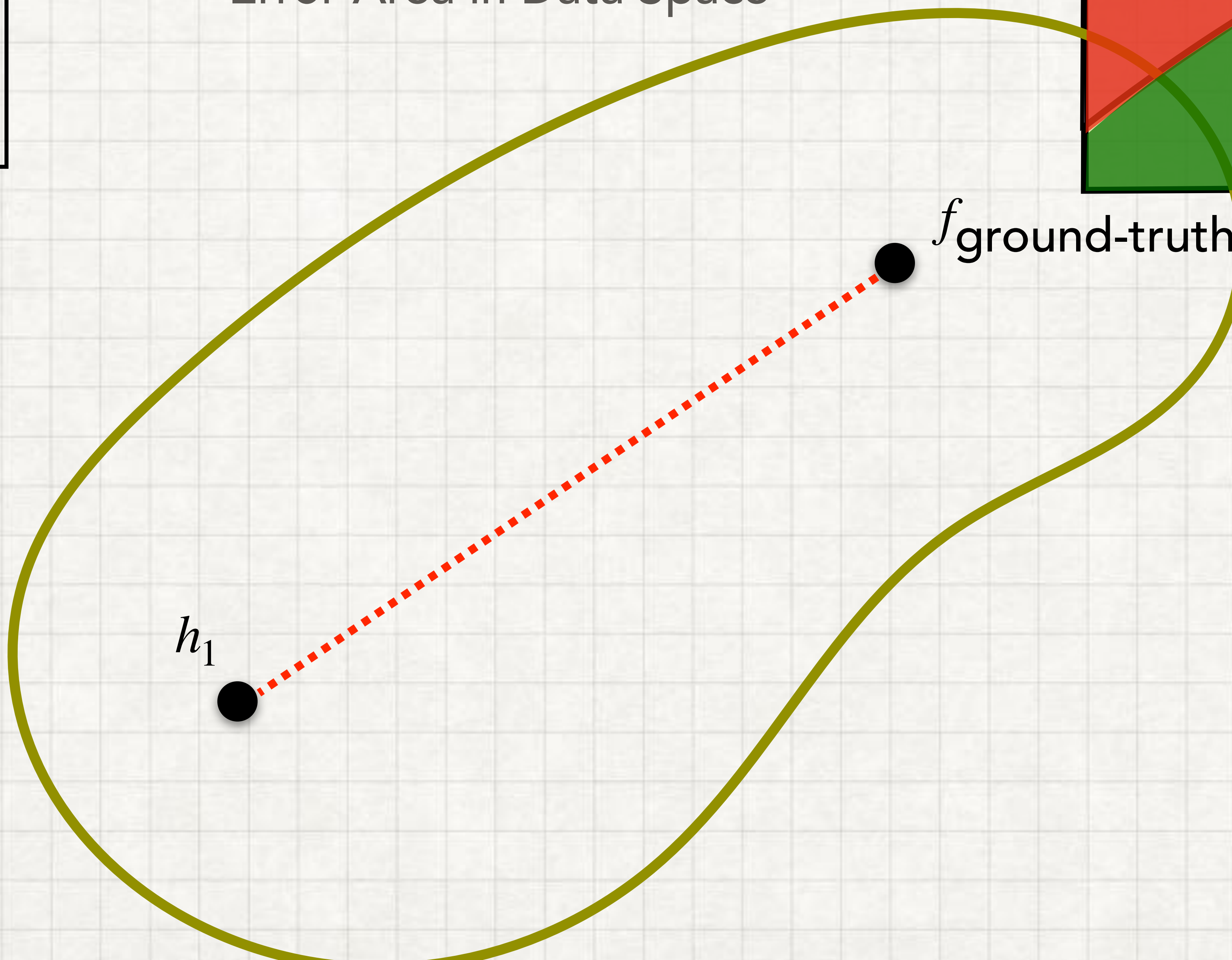
ERROR/RISK IN FUNCTIONAL SPACE

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

setosa /
versicolor ?

Risk: Distance in Functional Space
Error Area in Data Space





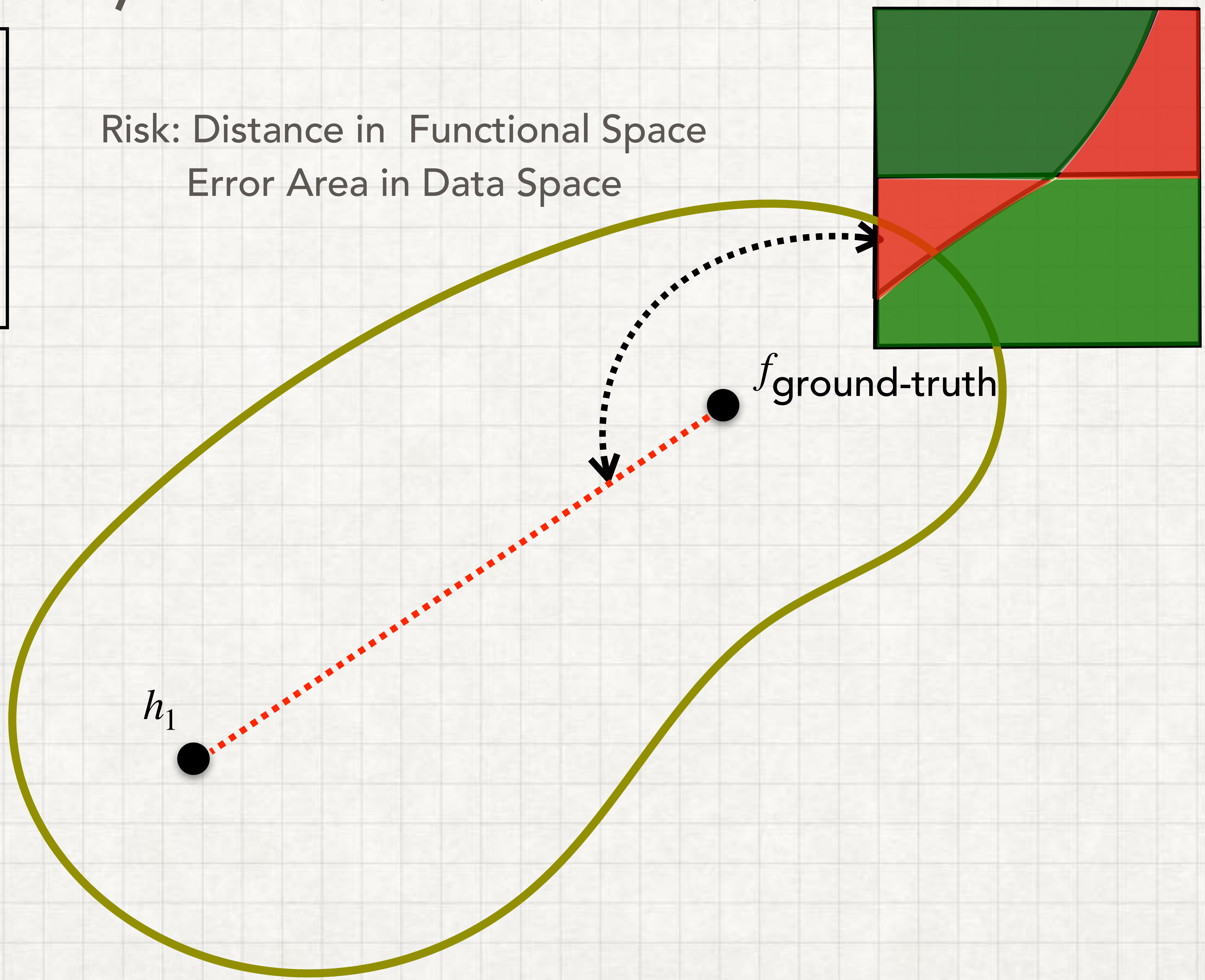
ERROR/RISK IN FUNCTIONAL SPACE

$\mathscr{X}, \mathscr{Y}$

setosa / versicolor ?

Risk: Distance in Functional Space

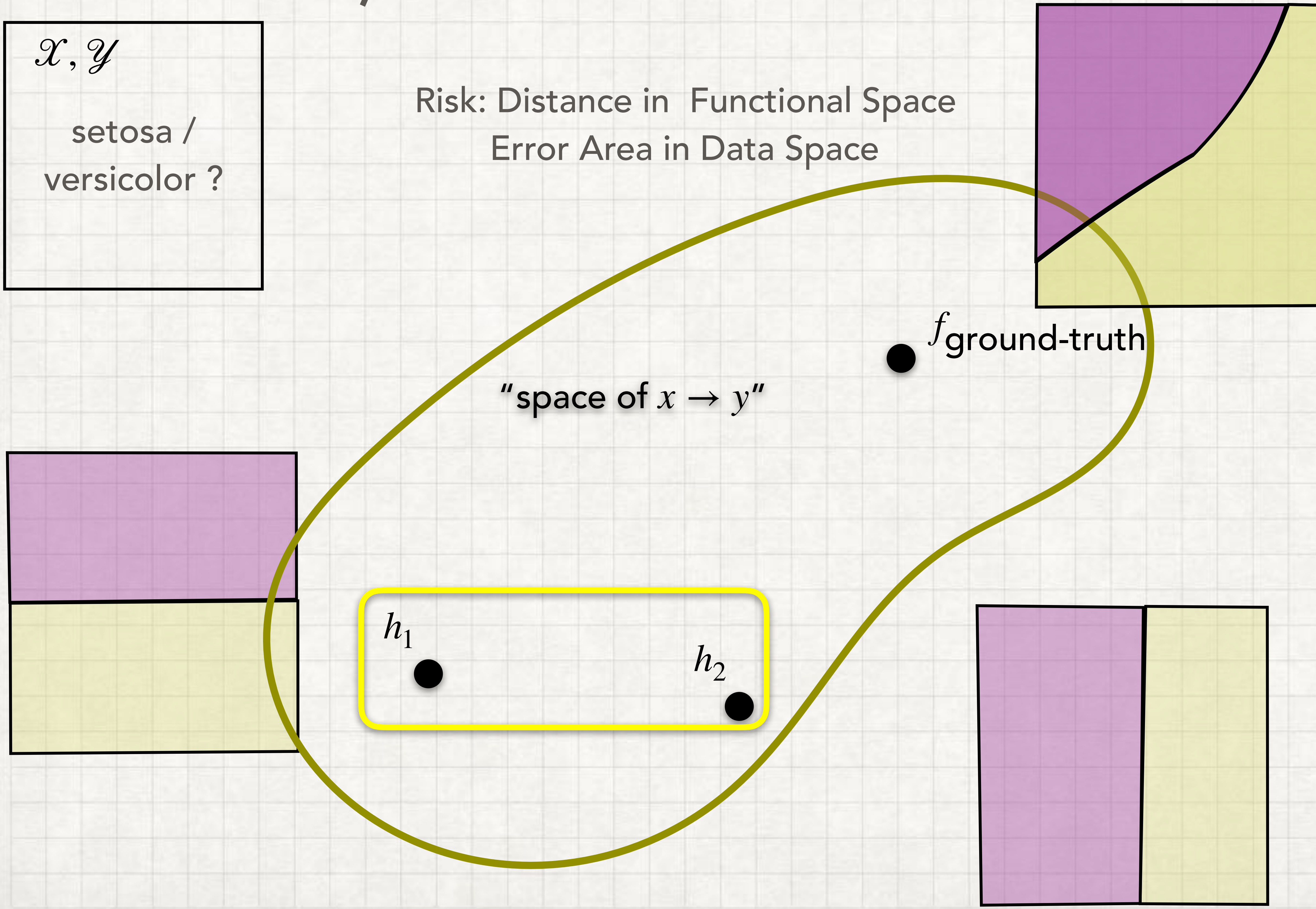
Error Area in Data Space



ERROR/RISKS OF TWO HYPOTHESES

Risk: Distance in Functional Space

Error Area in Data Space



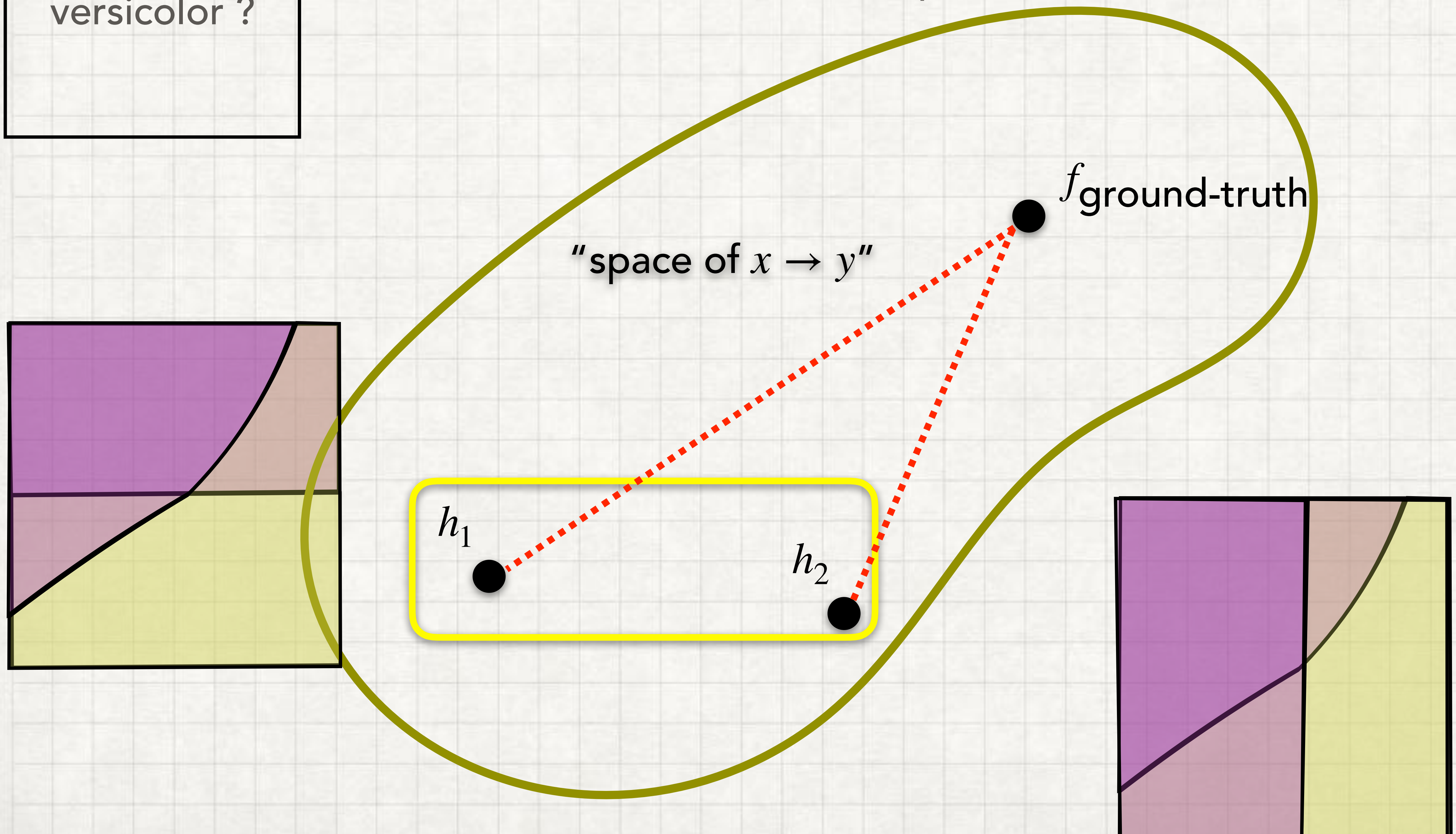
ERROR/RISKS OF TWO HYPOTHESES

$\mathscr{X}, \mathscr{Y}$

setosa / versicolor ?

Risk: Distance in Functional Space

Error Area in Data Space



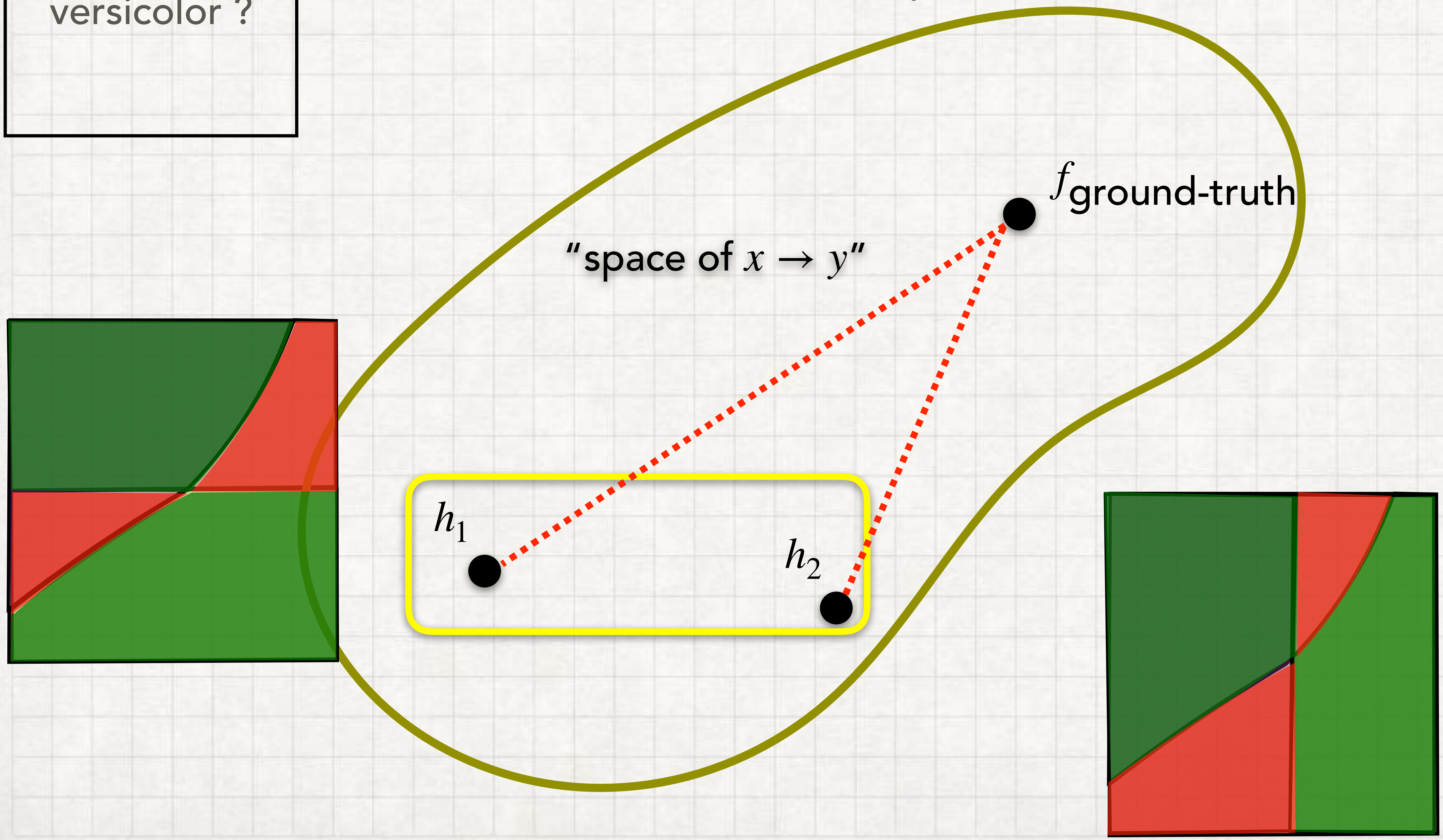
ERROR/RISKS OF TWO HYPOTHESES

$\mathscr{X}, \mathscr{Y}$

setosa / versicolor ?

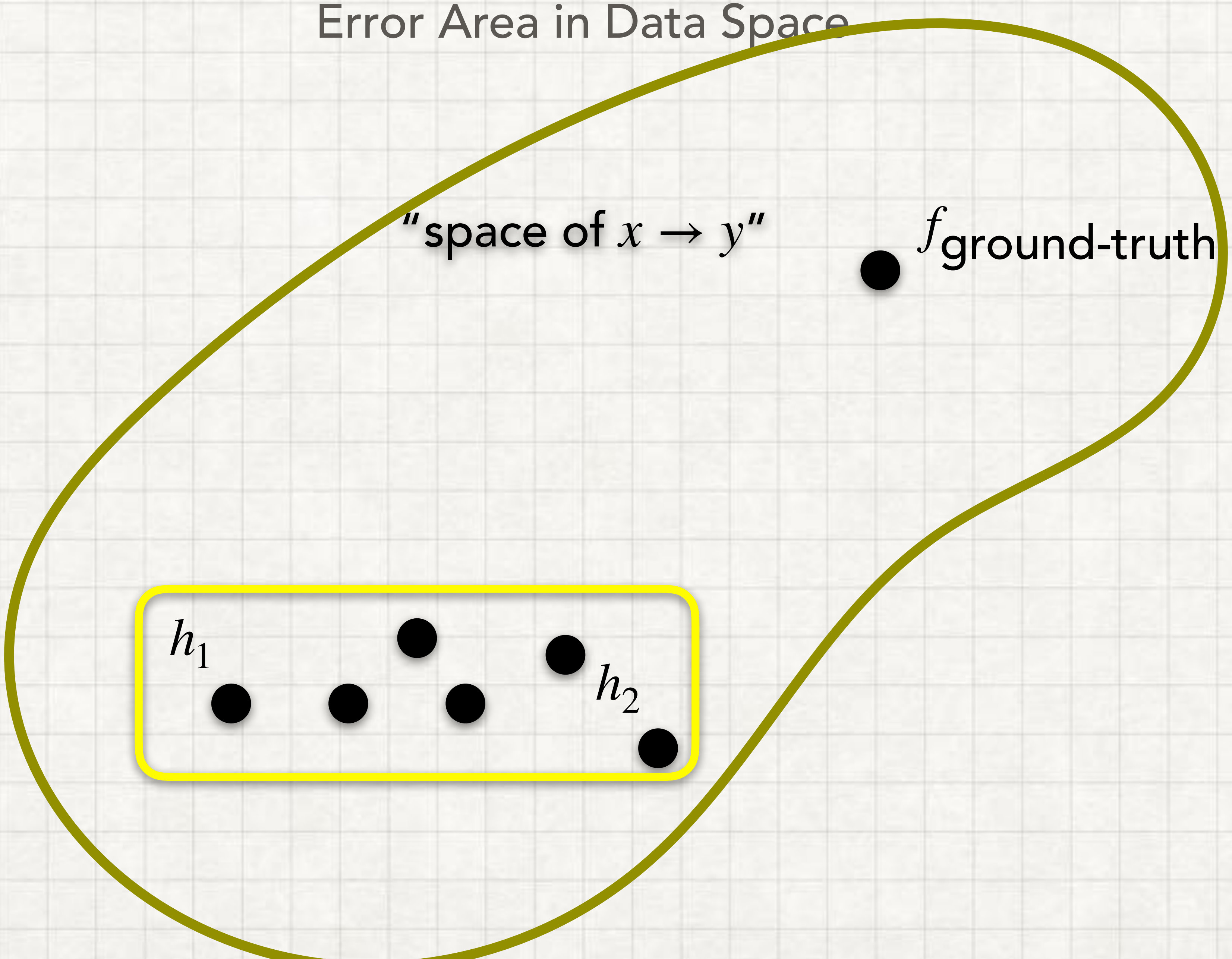
Risk: Distance in Functional Space

Error Area in Data Space



ERROR/RISKS OF MANY HYPOTHESES

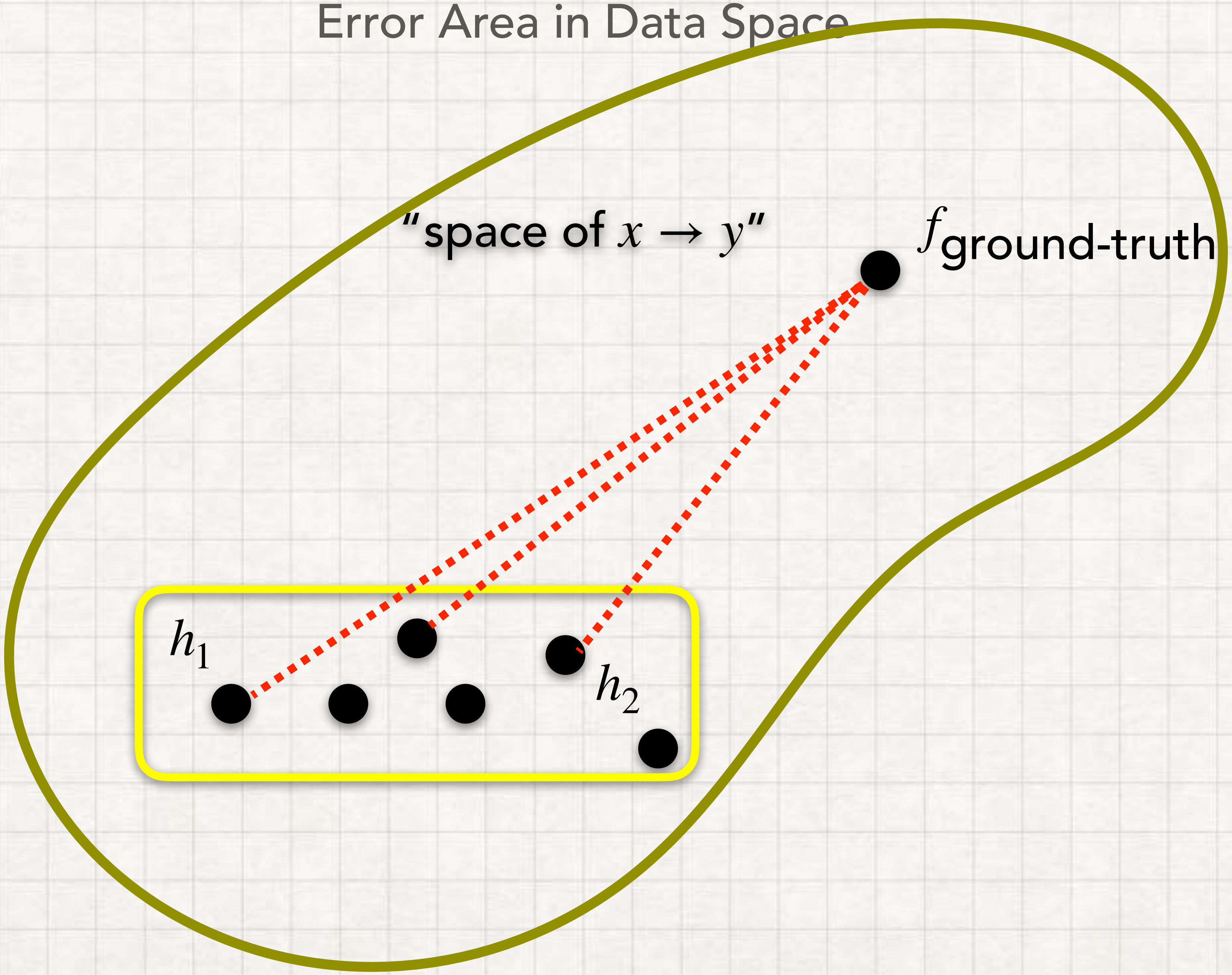
Risk: Distance in Functional Space
Error Area in Data Space



ERROR/RISKS OF MANY HYPOTHESES

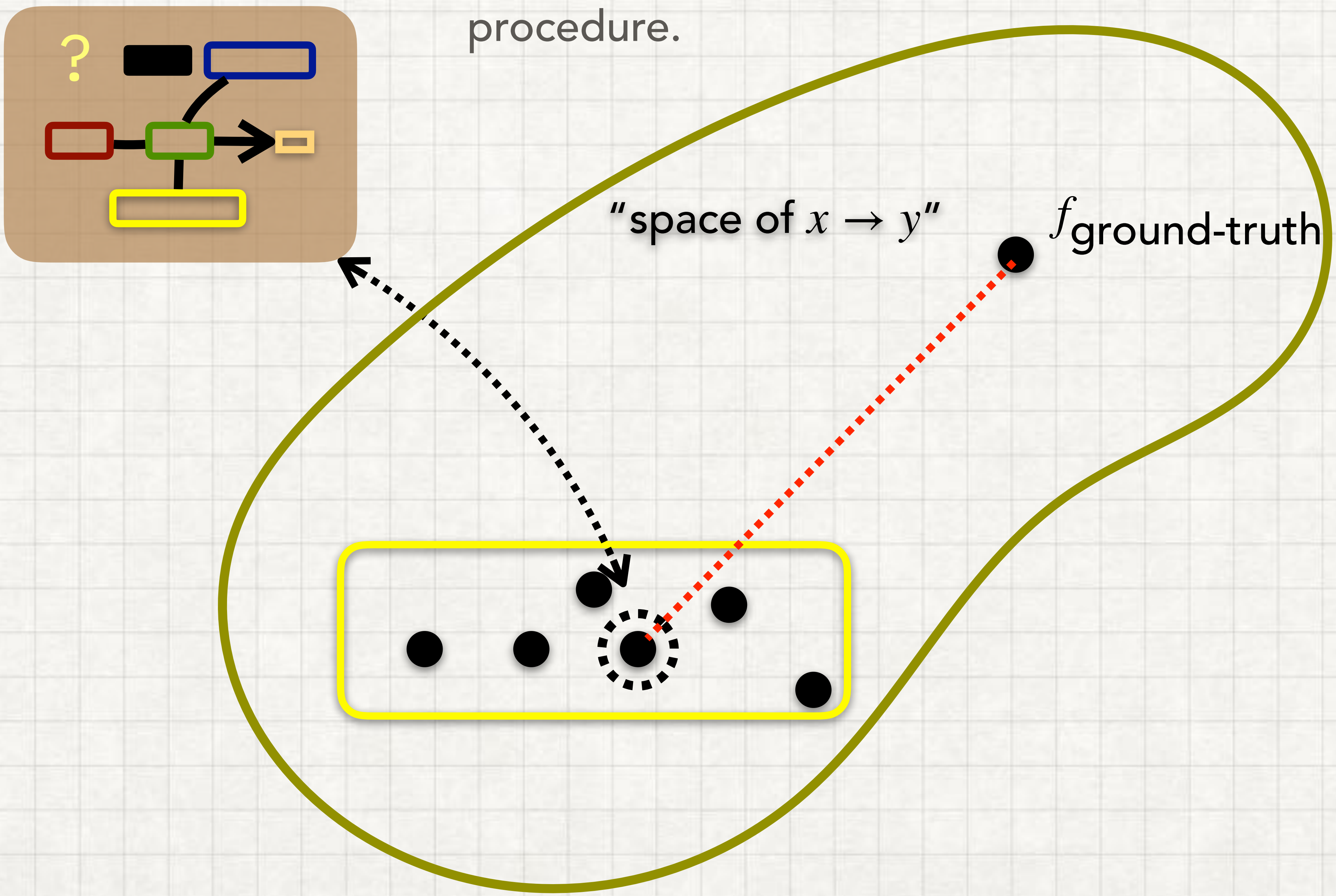
Risk: Distance in Functional Space

Error Area in Data Space



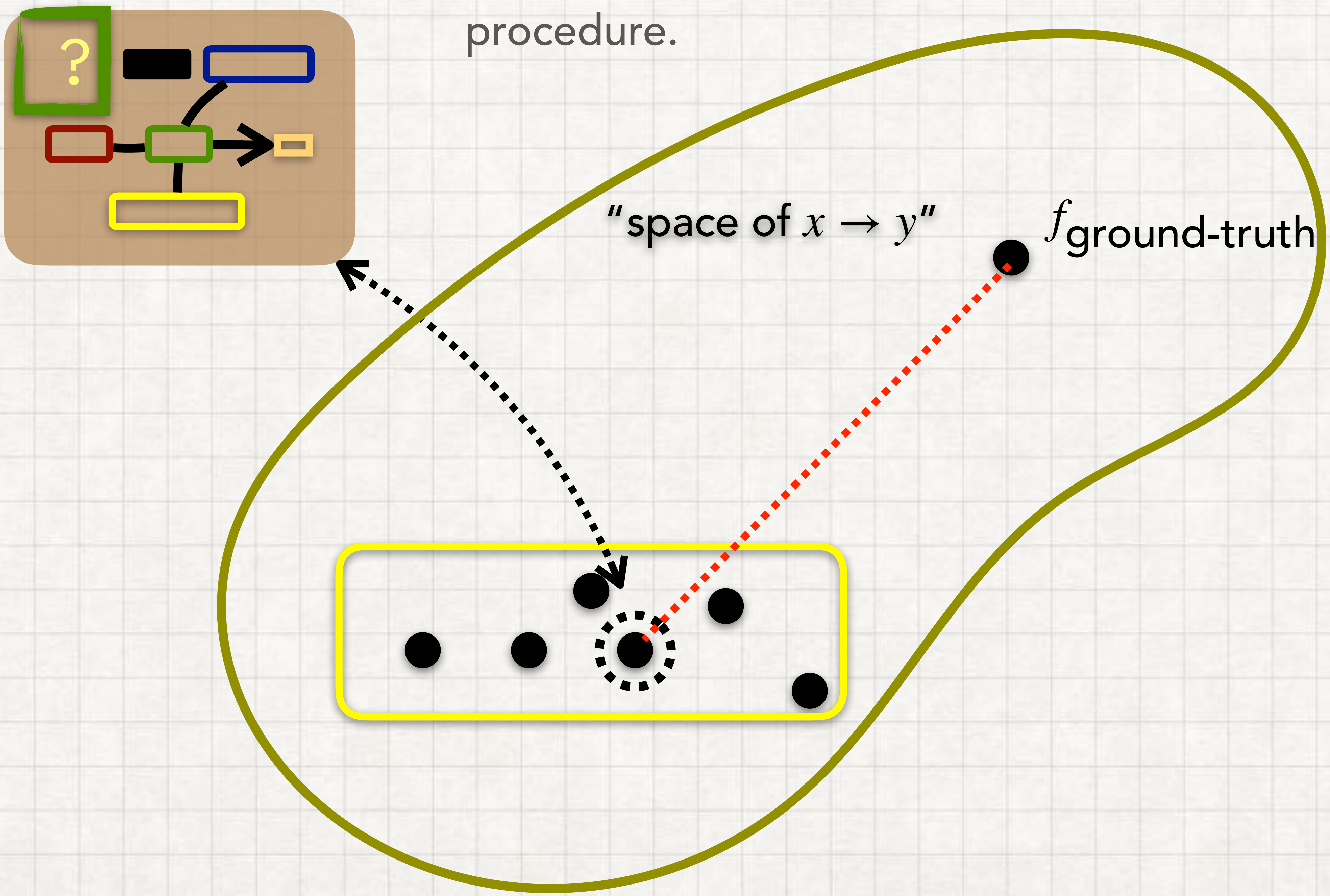
ERROR/RISKS OF SELECTED HYPOTHESES FROM MANY

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

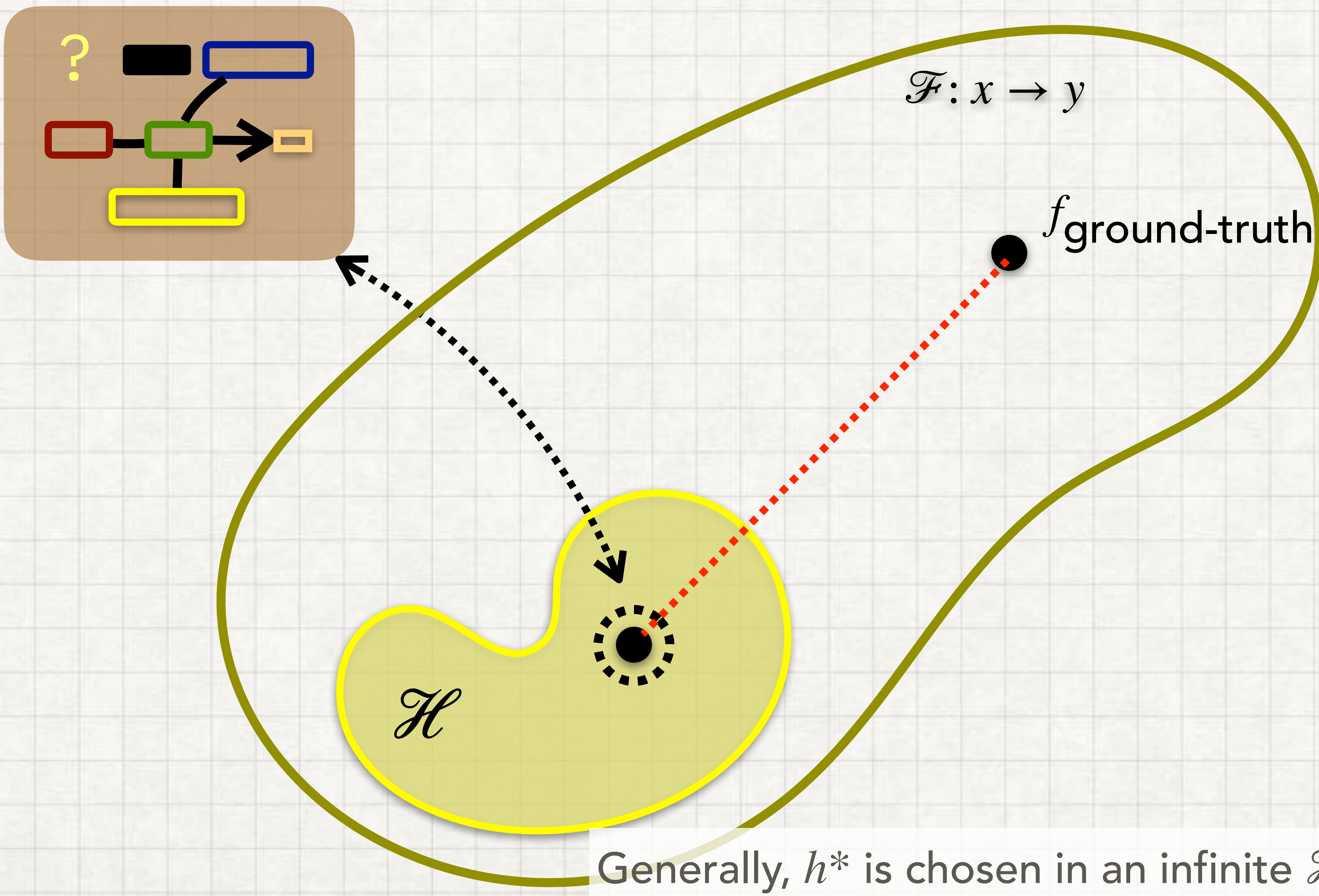


ERROR/RISKS OF SELECTED HYPOTHESES FROM MANY

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



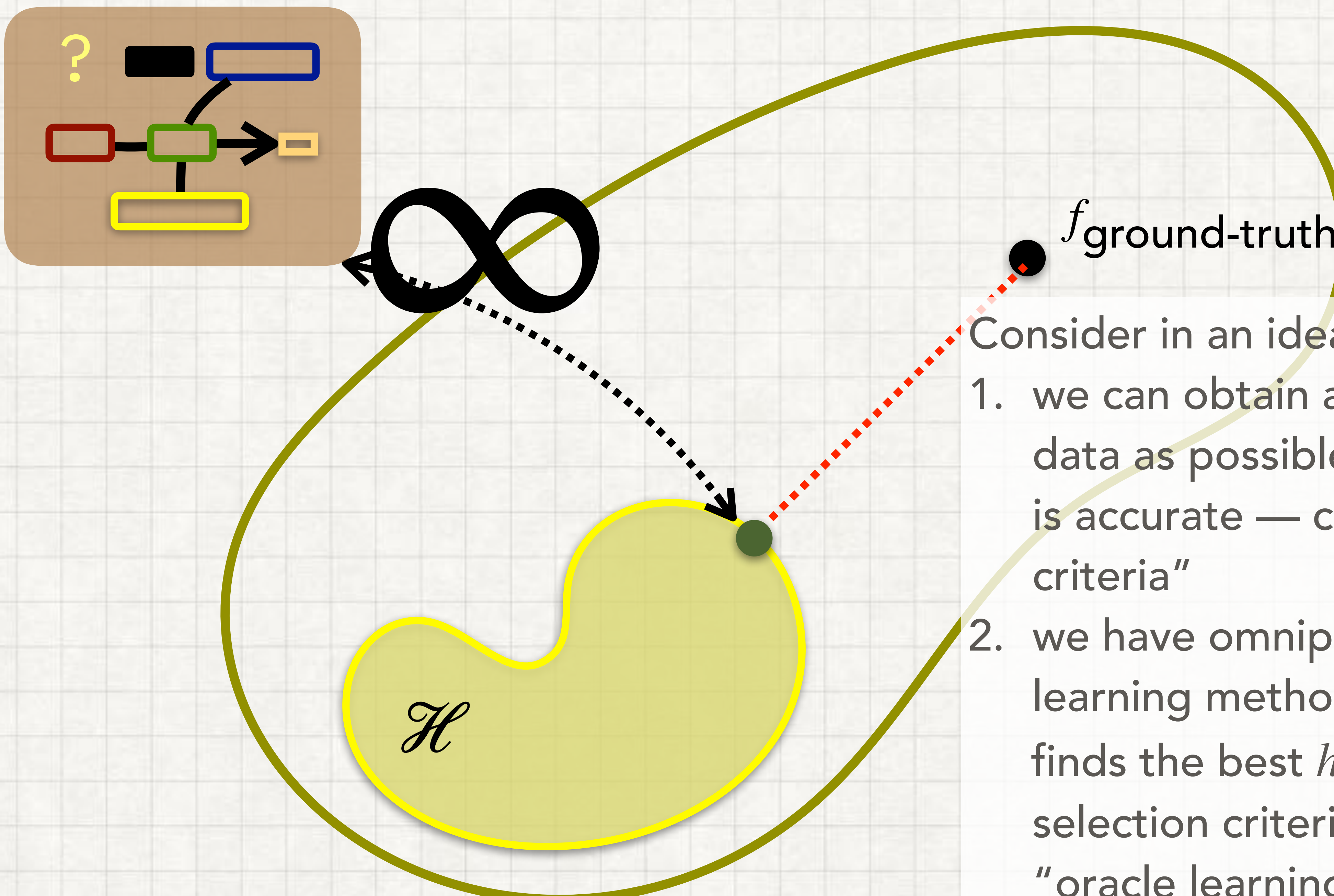
ERROR/RISKS OF SELECTED HYPOTHESIS FROM $\mathscr{H}$



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"

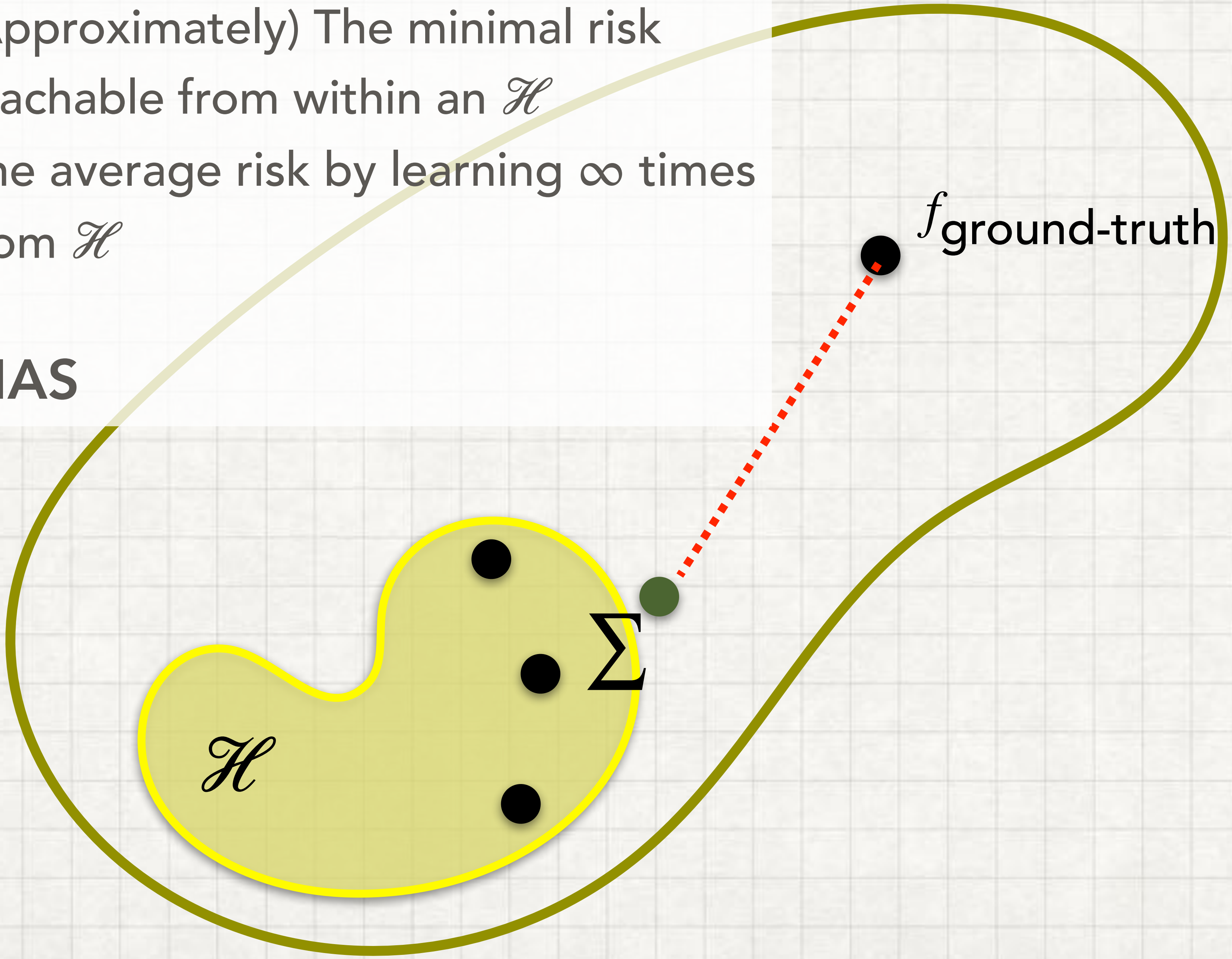
Putting together, we then consider the minimal risk of an $\mathscr{H}$.

DEFINITION OF BIAS

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

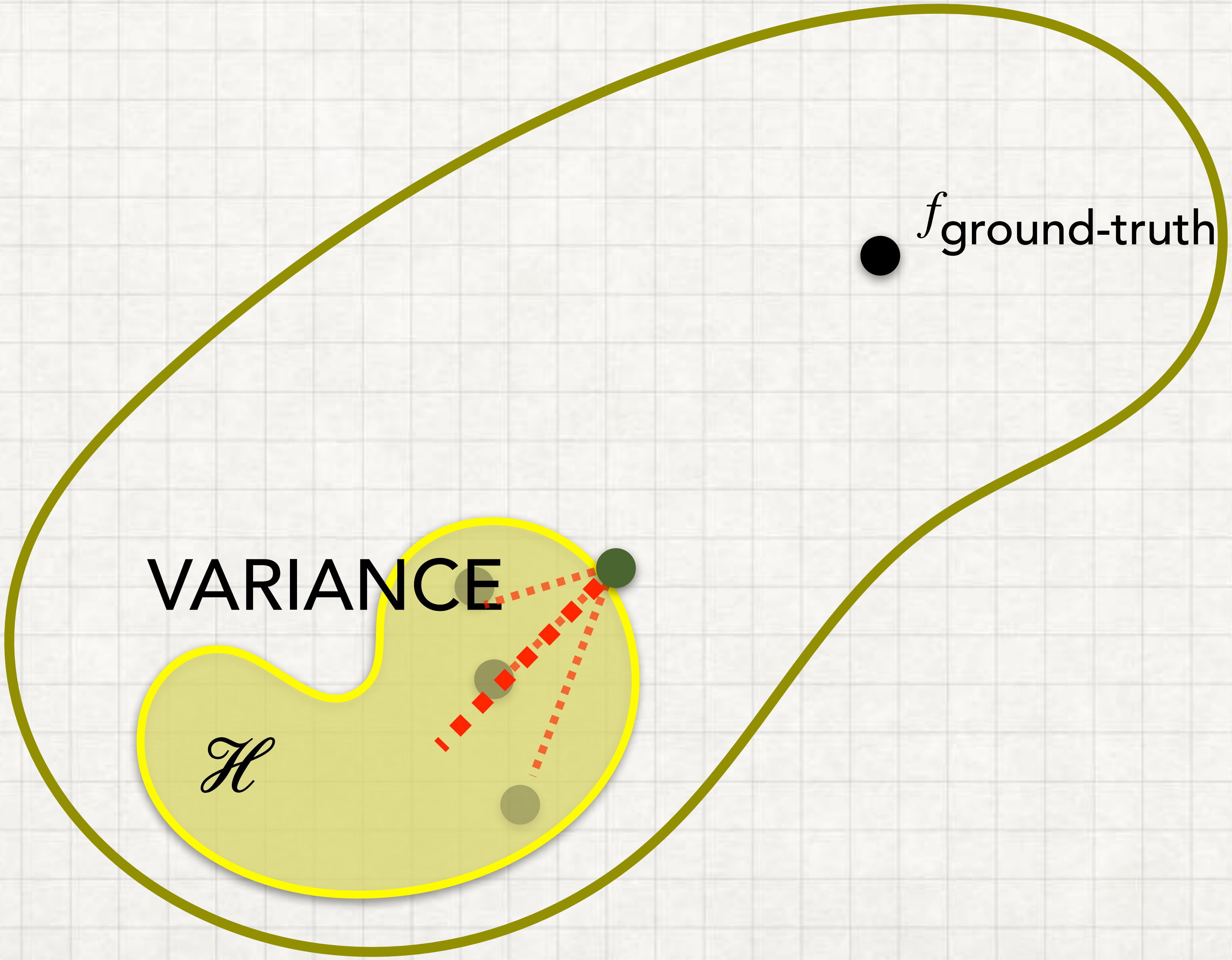
The average risk by learning ∞ times from $\mathscr{H}$

∞ **BIAS**



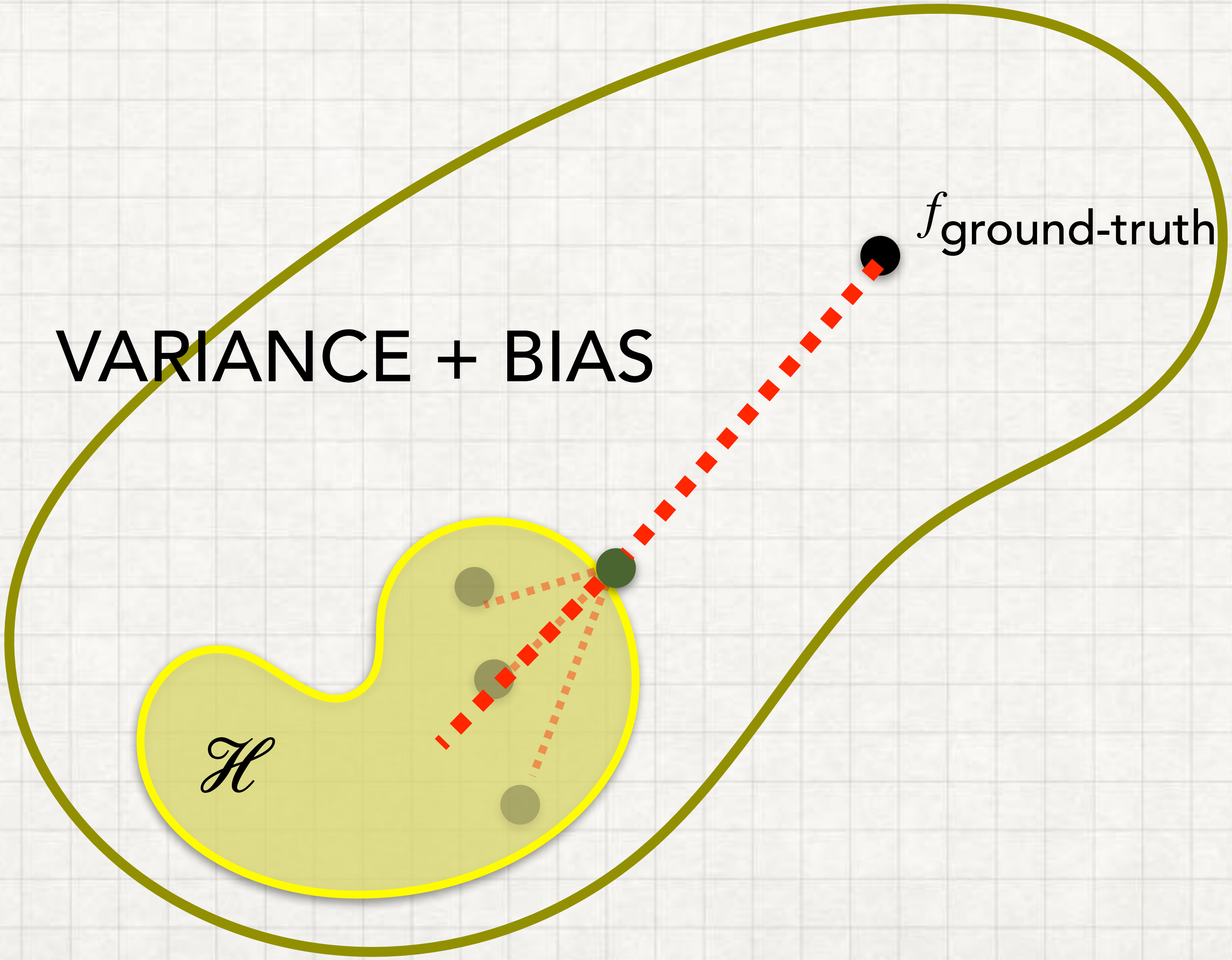
RISK OF TRAINING WITH ONE, RANDOM

D_{train}



TOTAL RISK OF A TRAINING

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

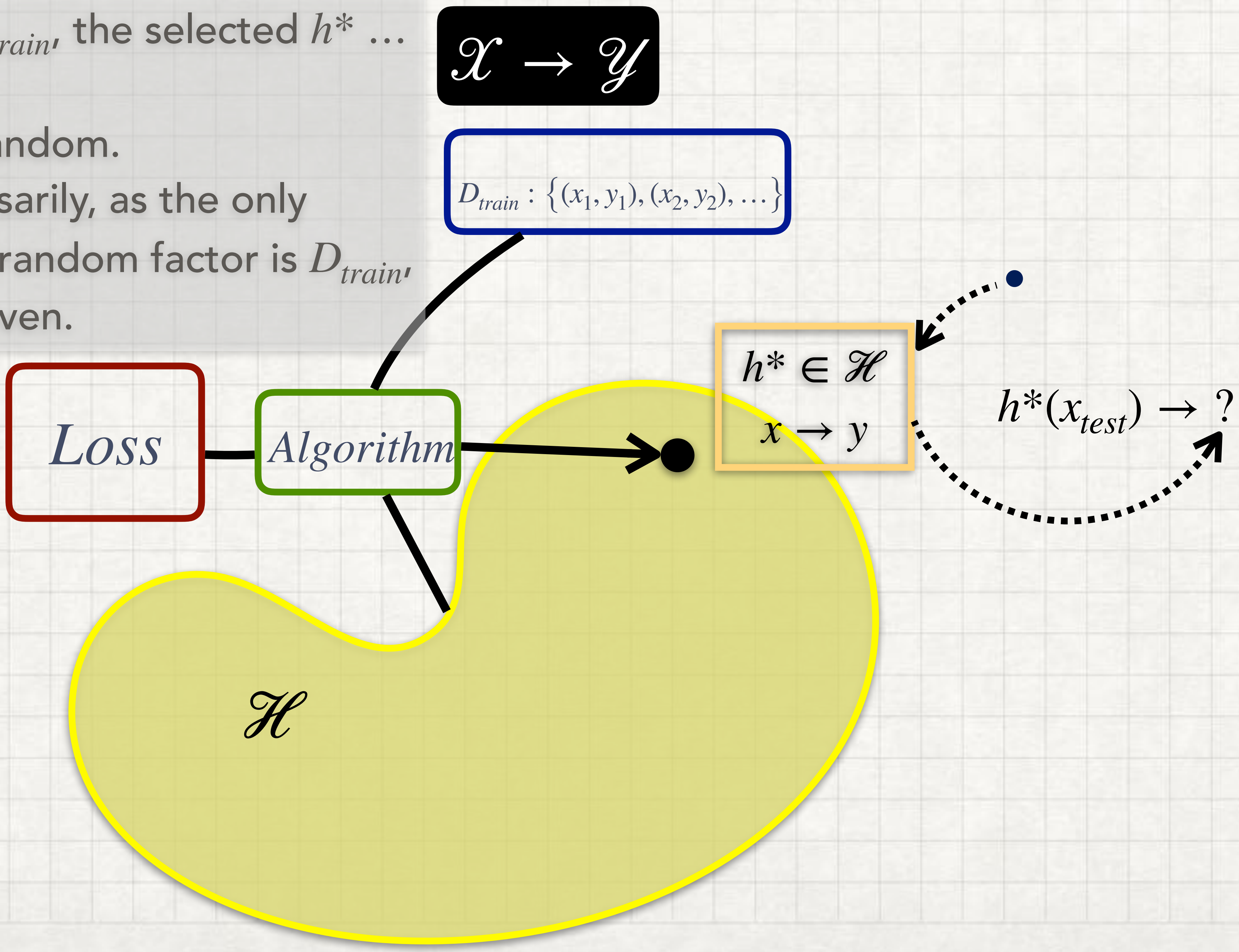


SUMMARISE: RANDOMISATION OF THE PREDICTION

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

Q: Given a D_{train}, the selected h^* …

A. must be random.
B. Not necessarily, as the only inevitable random factor is D_{train}, which is given.

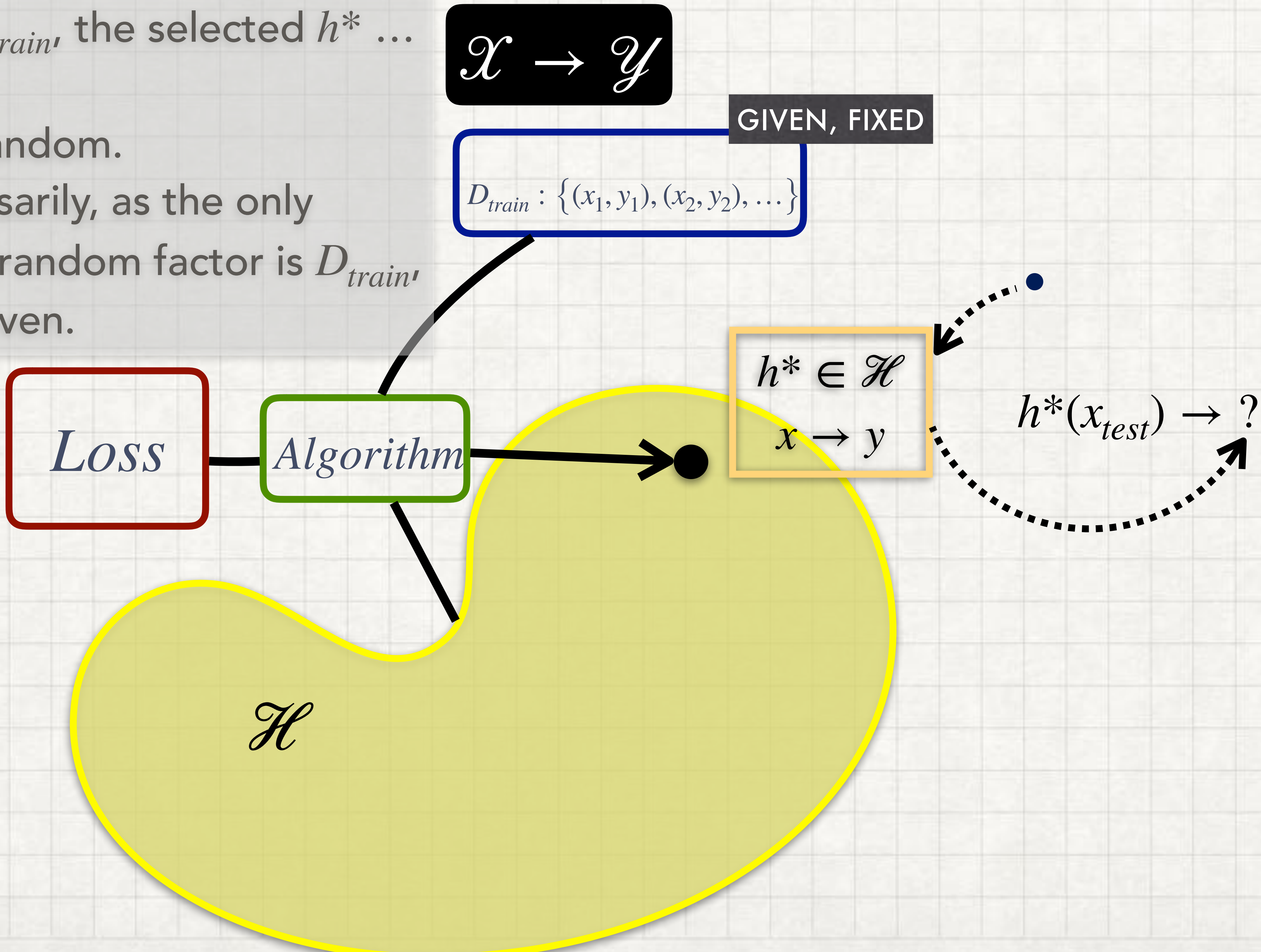


SUMMARISE: RANDOMISATION OF THE PREDICTION

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

Q: Given a D_{train}, the selected h^* …

A. must be random.
B. Not necessarily, as the only inevitable random factor is D_{train}, which is given.



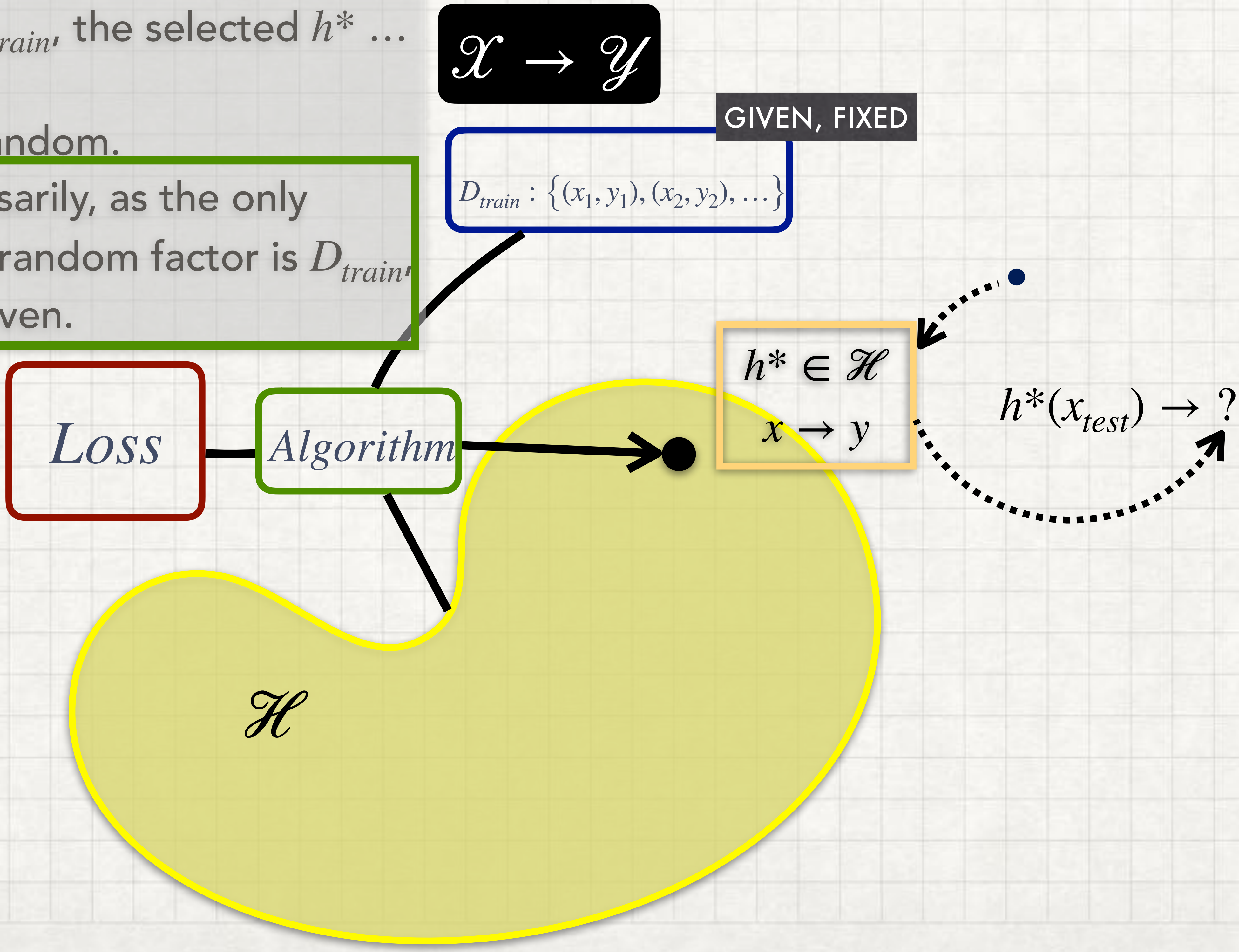
SUMMARISE: RANDOMISATION OF THE PREDICTION

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

Q: Given a D_{train}, the selected h^* …

A. must be random.

B. Not necessarily, as the only inevitable random factor is D_{train}, which is given.



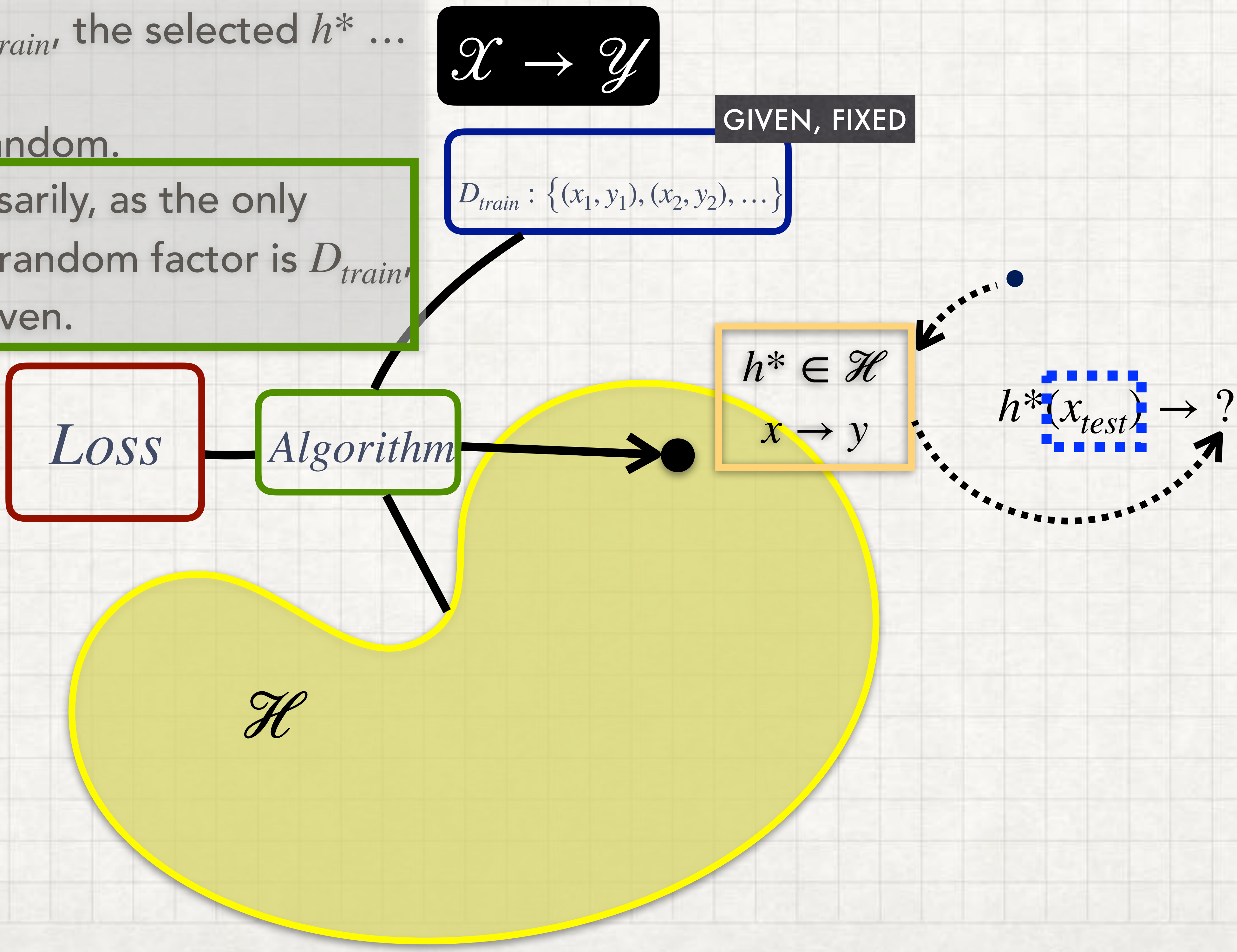
SUMMARISE: RANDOMISATION OF THE PREDICTION

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

Q: Given a D_{train}, the selected h^* …

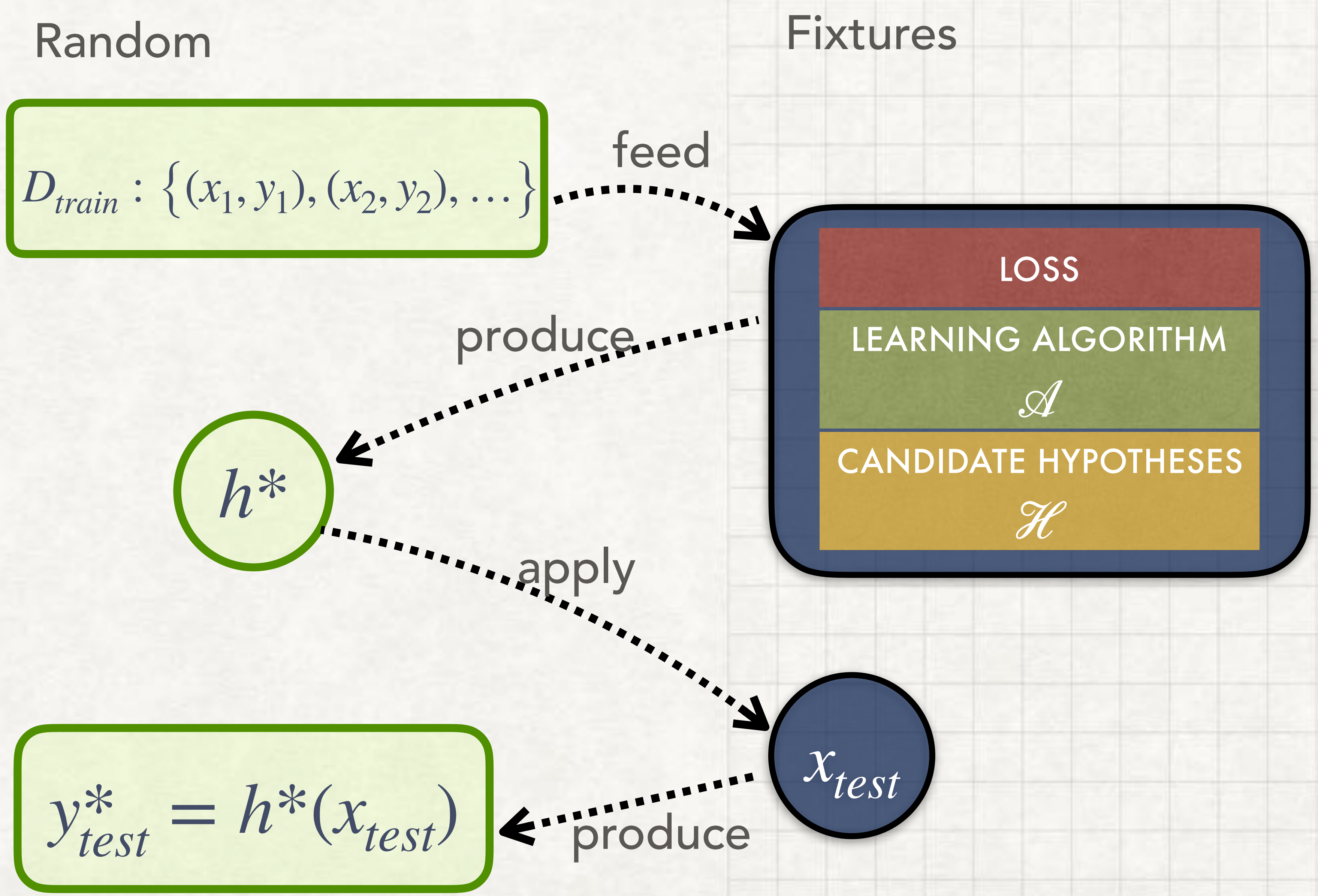
A. must be random.

B. Not necessarily, as the only inevitable random factor is D_{train}, which is given.



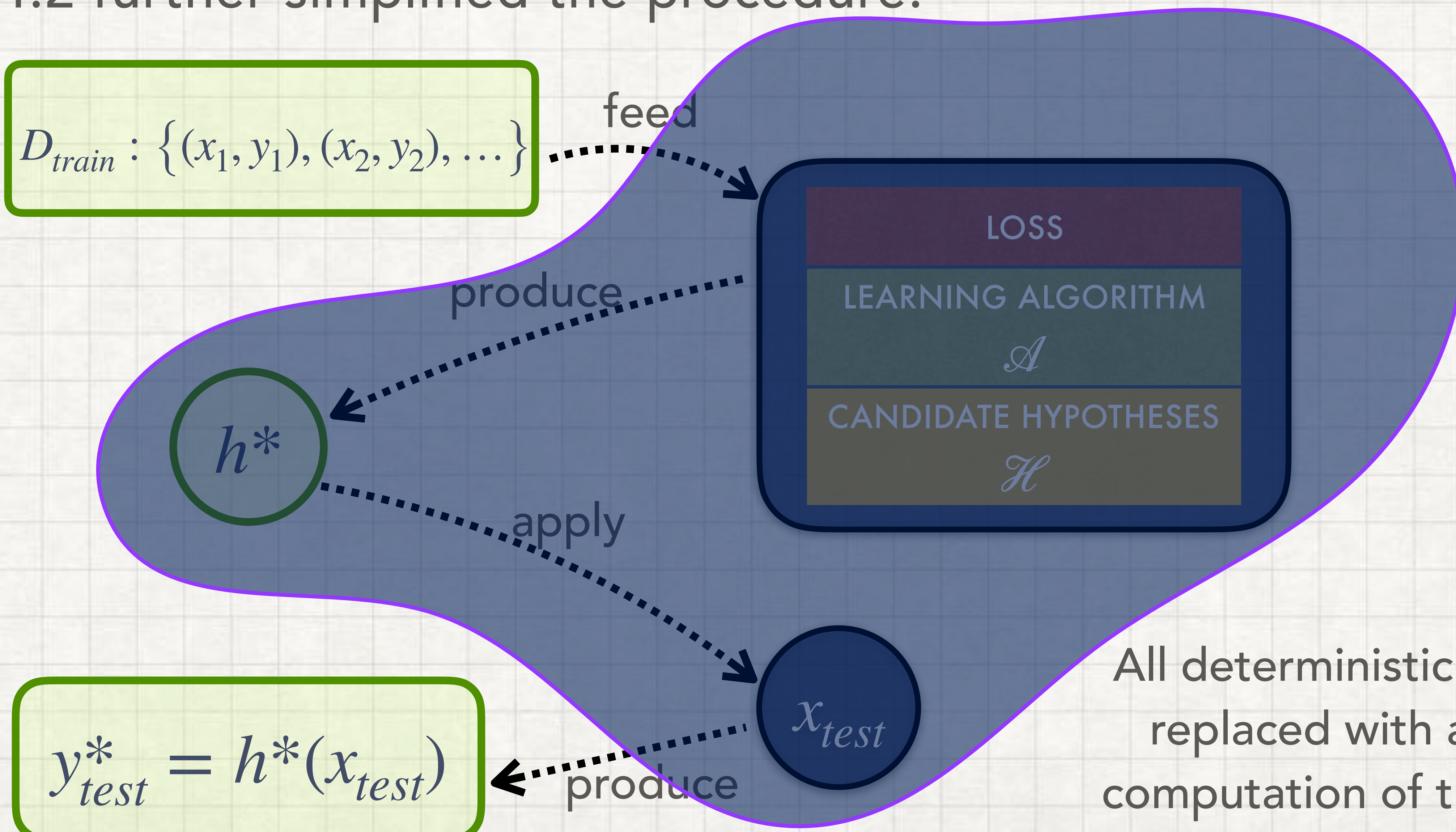
RANDOMISATION ANALYSIS

- 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:



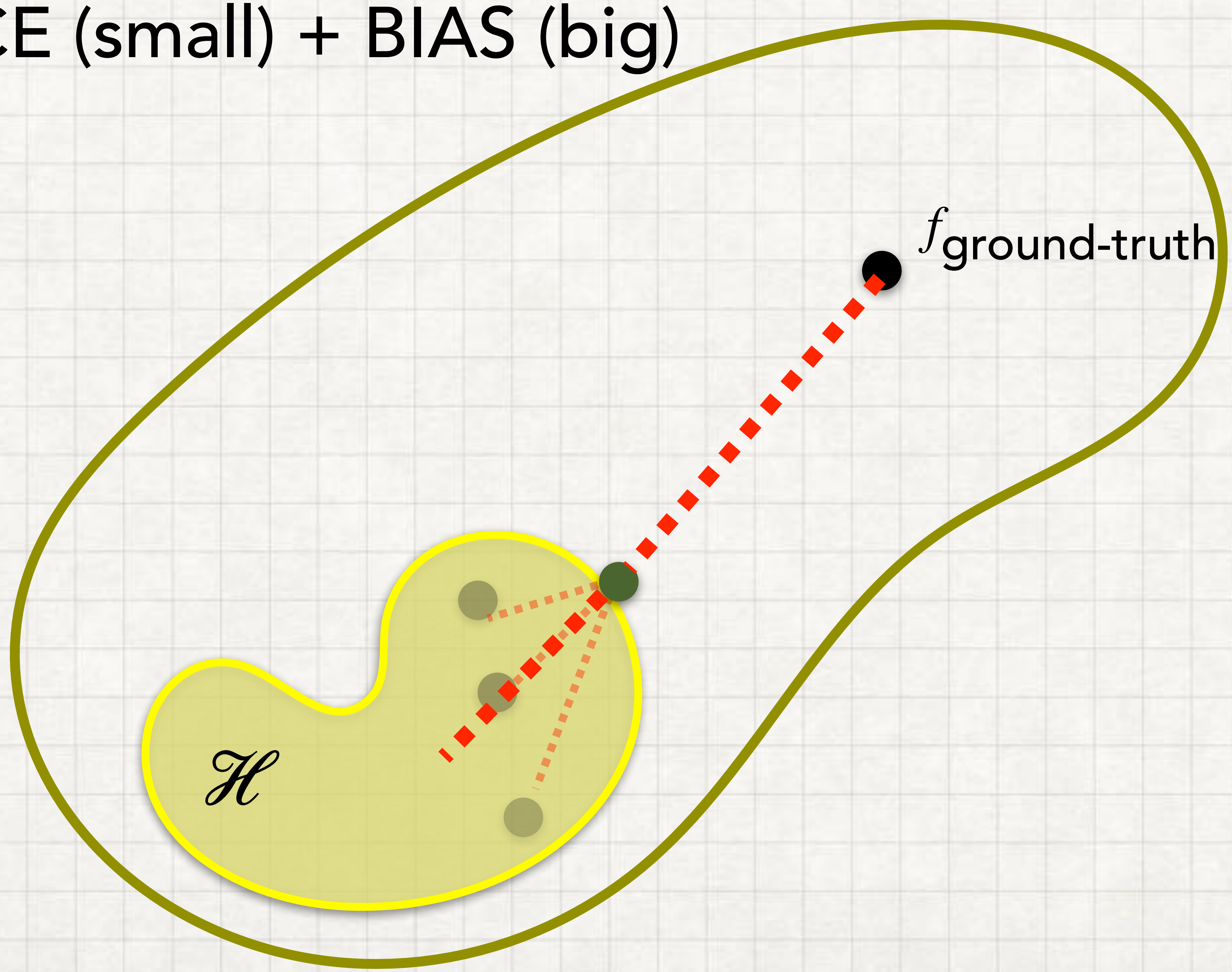
All deterministic steps are replaced with a simple computation of the sample median.

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)

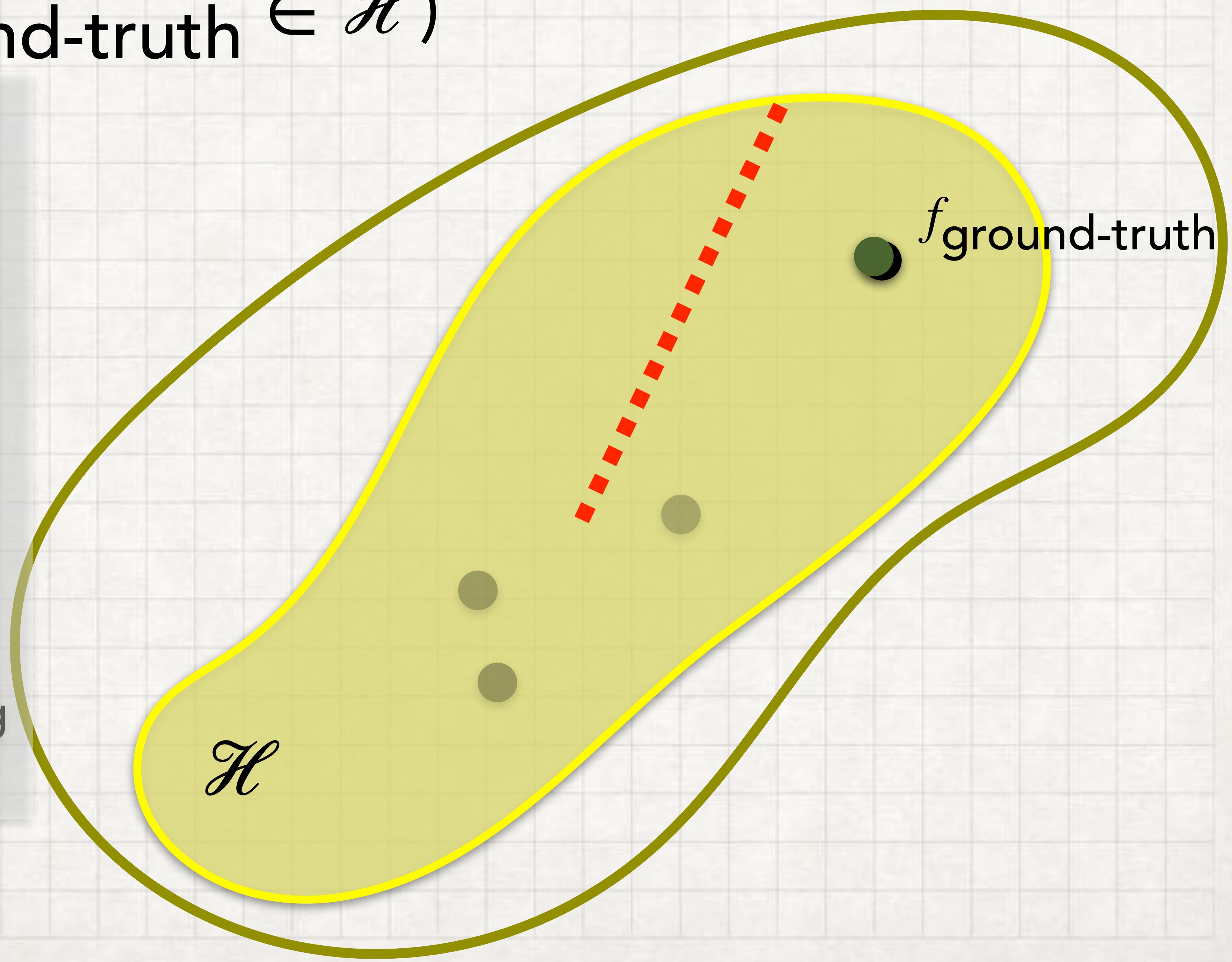


TOTAL RISK OF A TRAINING
BIG $\mathscr{H}$

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

Q: Given $\mathscr{H}$, can we reduce the variance?

A. No, variance is an intrinsic property of $\mathscr{H}$.
B. Yes, by using less training samples, so the training result will not change as much.
C. Yes, by using more training samples.

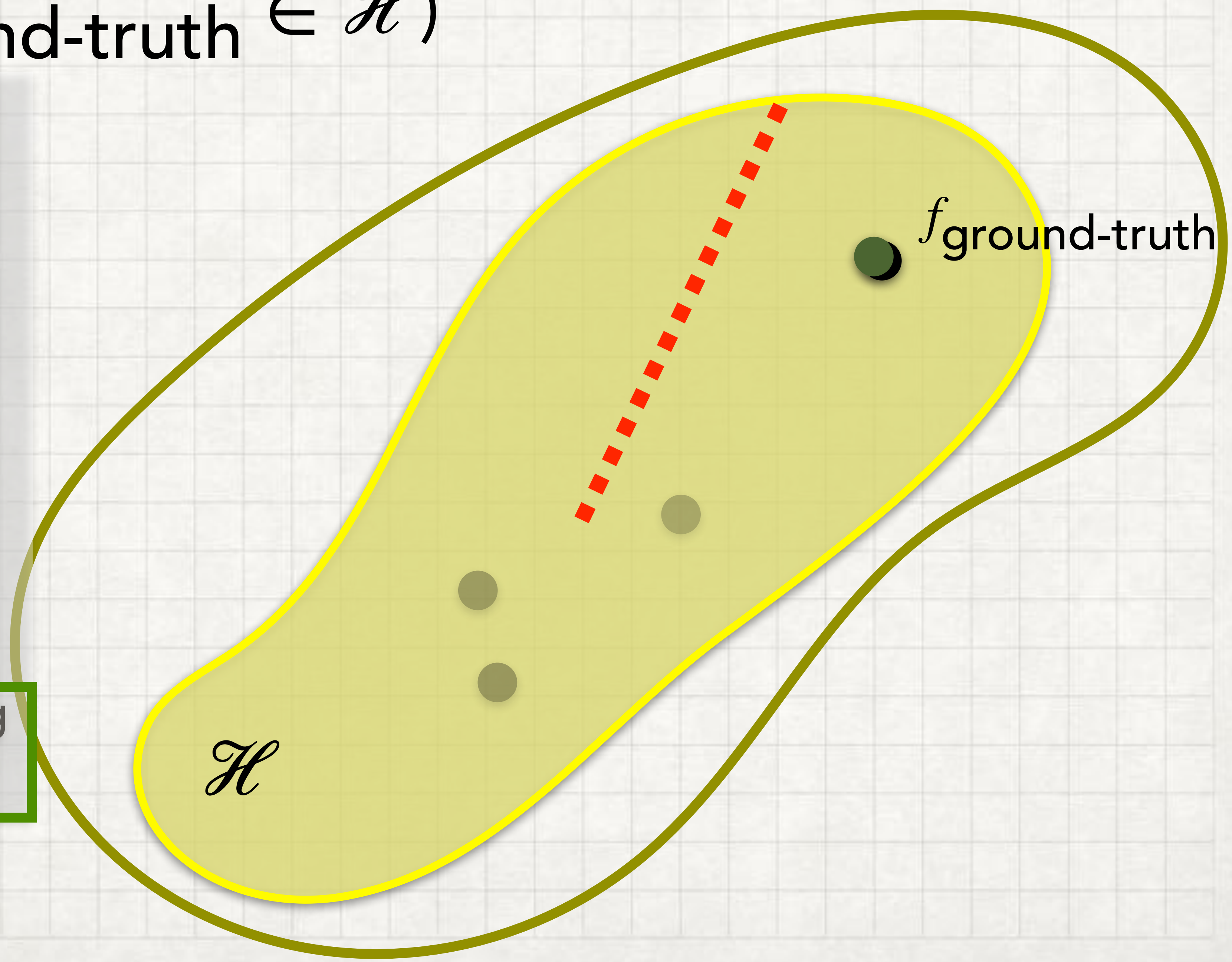


TOTAL RISK OF A TRAINING BIG $\mathscr{H}$

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

Q: Given $\mathscr{H}$, can we reduce the variance?

A. No, variance is an intrinsic property of $\mathscr{H}$.
B. Yes, by using less training samples, so the training result will not change as much.
C. Yes, by using more training samples.

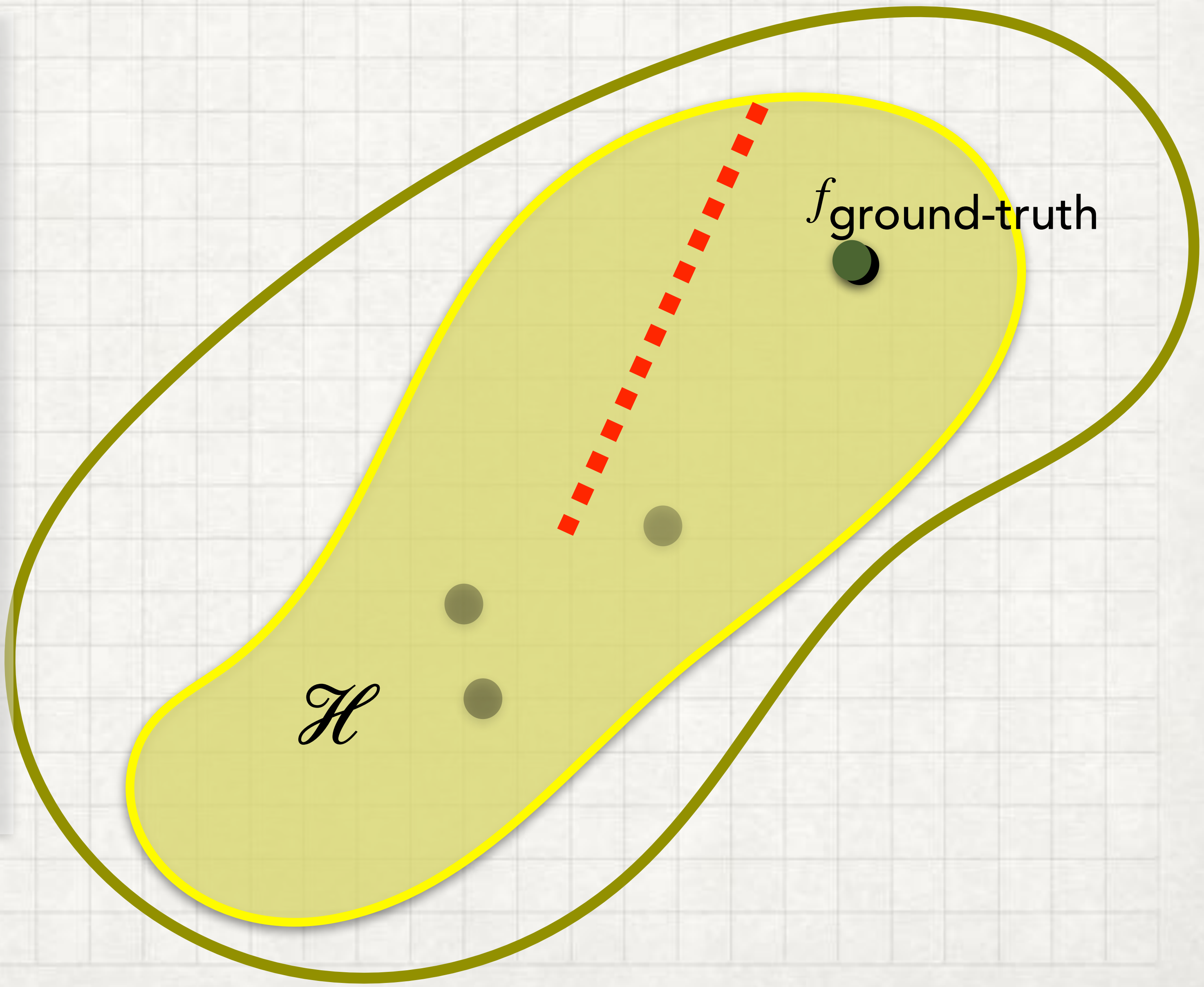


TOTAL RISK OF A TRAINING
BIG $\mathscr{H}$

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

Q: Given $\mathscr{H}$, can we reduce the bias?

A. Very little, bias is an asymptotical property of $\mathscr{H}$, can not be easily changed by providing more training samples.
B. Yes, by using less training samples, so the training result will not change as much.
C. Yes, by using more training samples.

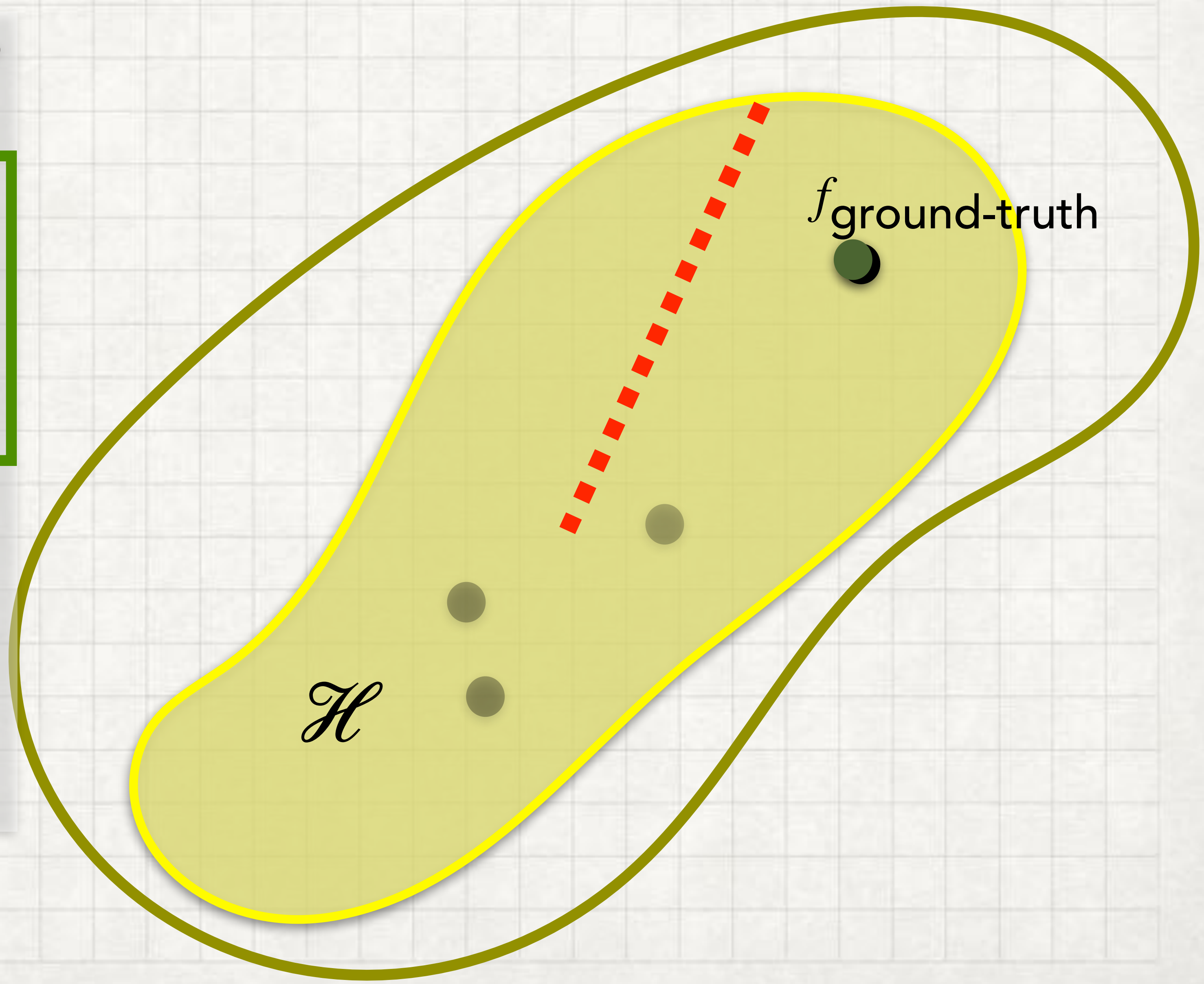


TOTAL RISK OF A TRAINING BIG $\mathscr{H}$

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

Q: Given $\mathscr{H}$, can we reduce the bias?

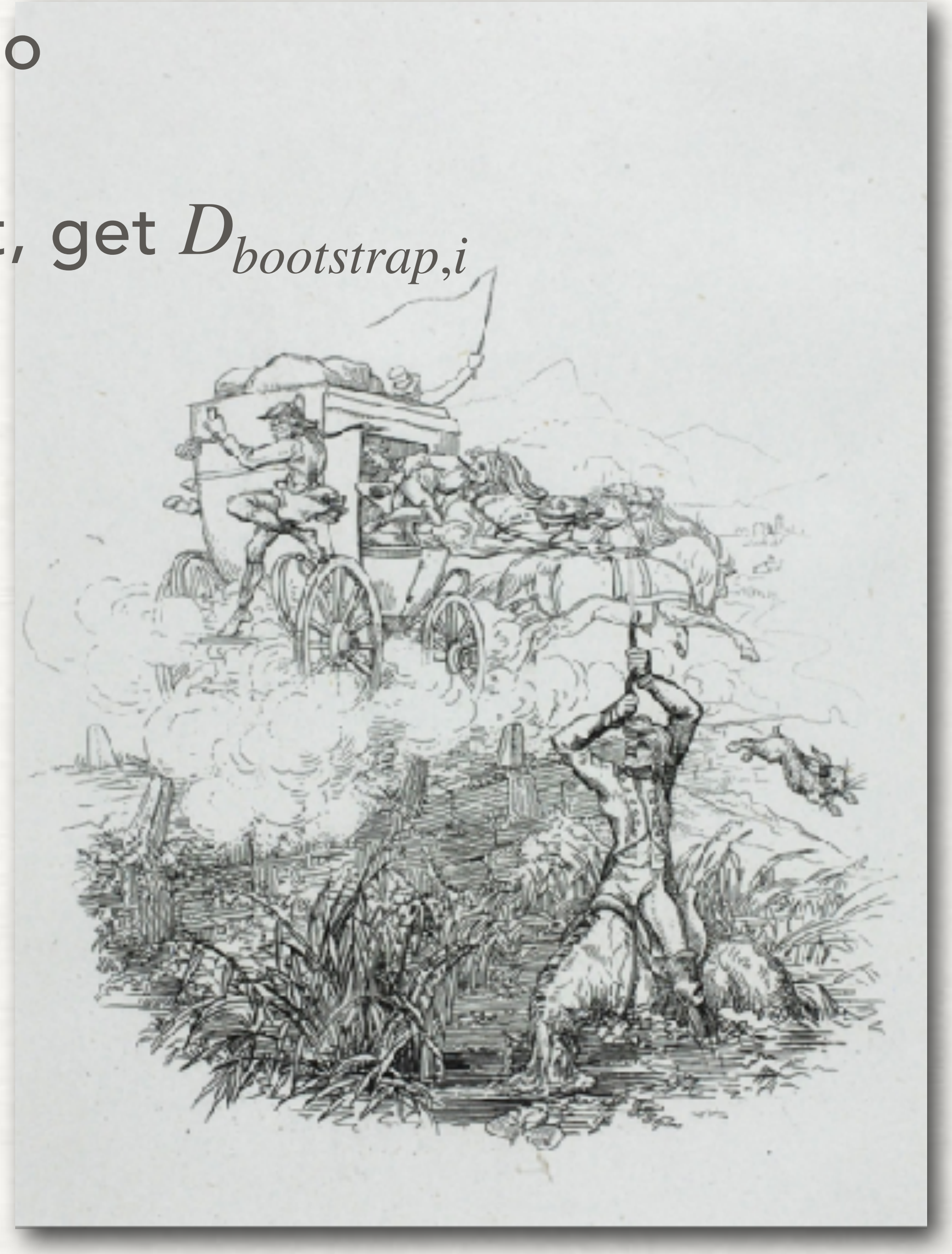
A. Very little, bias is an asymptotical property of $\mathscr{H}$, can not be easily changed by providing more training samples.
B. Yes, by using less training samples, so the training result will not change as much.
C. Yes, by using more training samples.



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

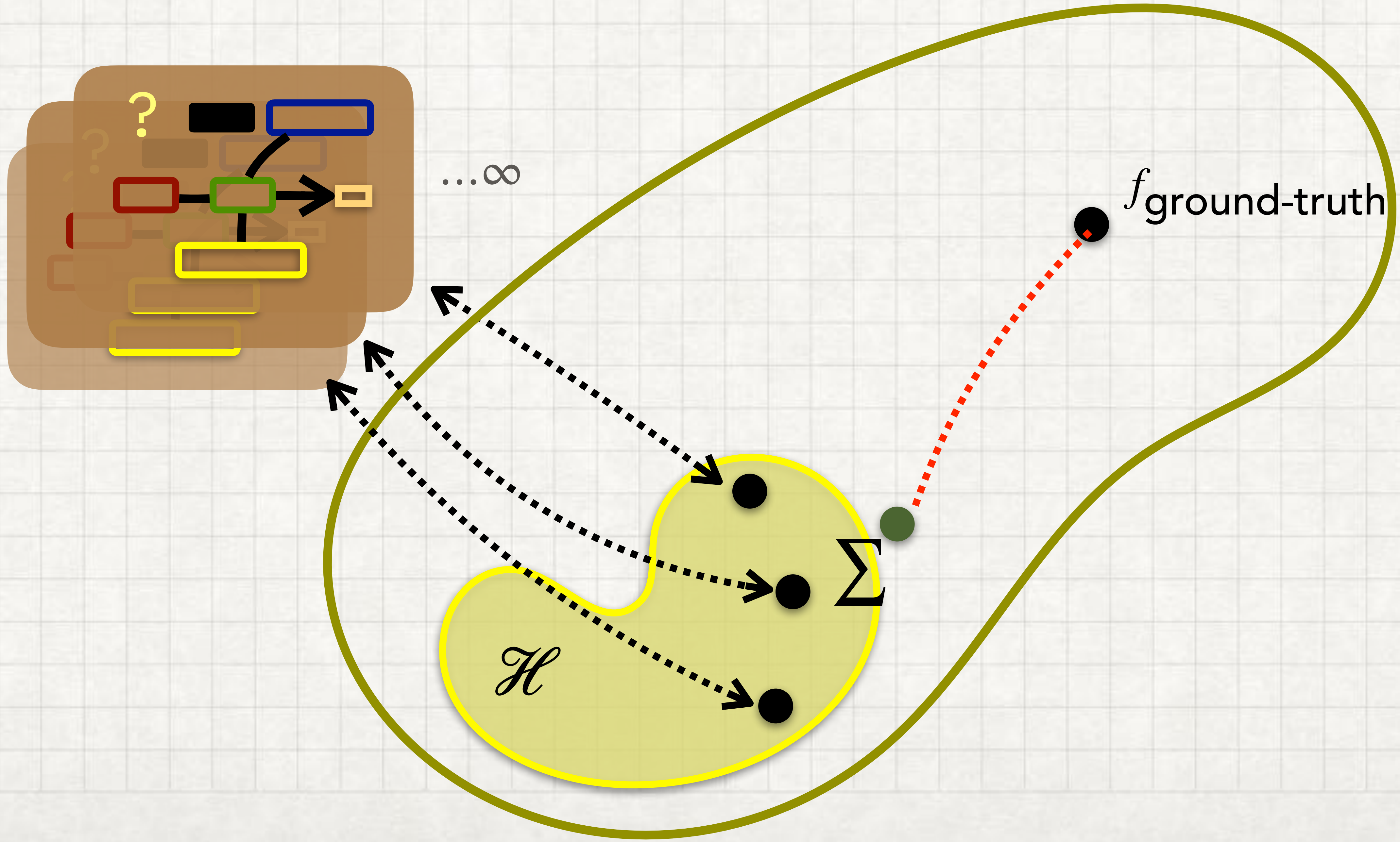
REVIEW: BOOTSTRAP / BAGGING

- 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



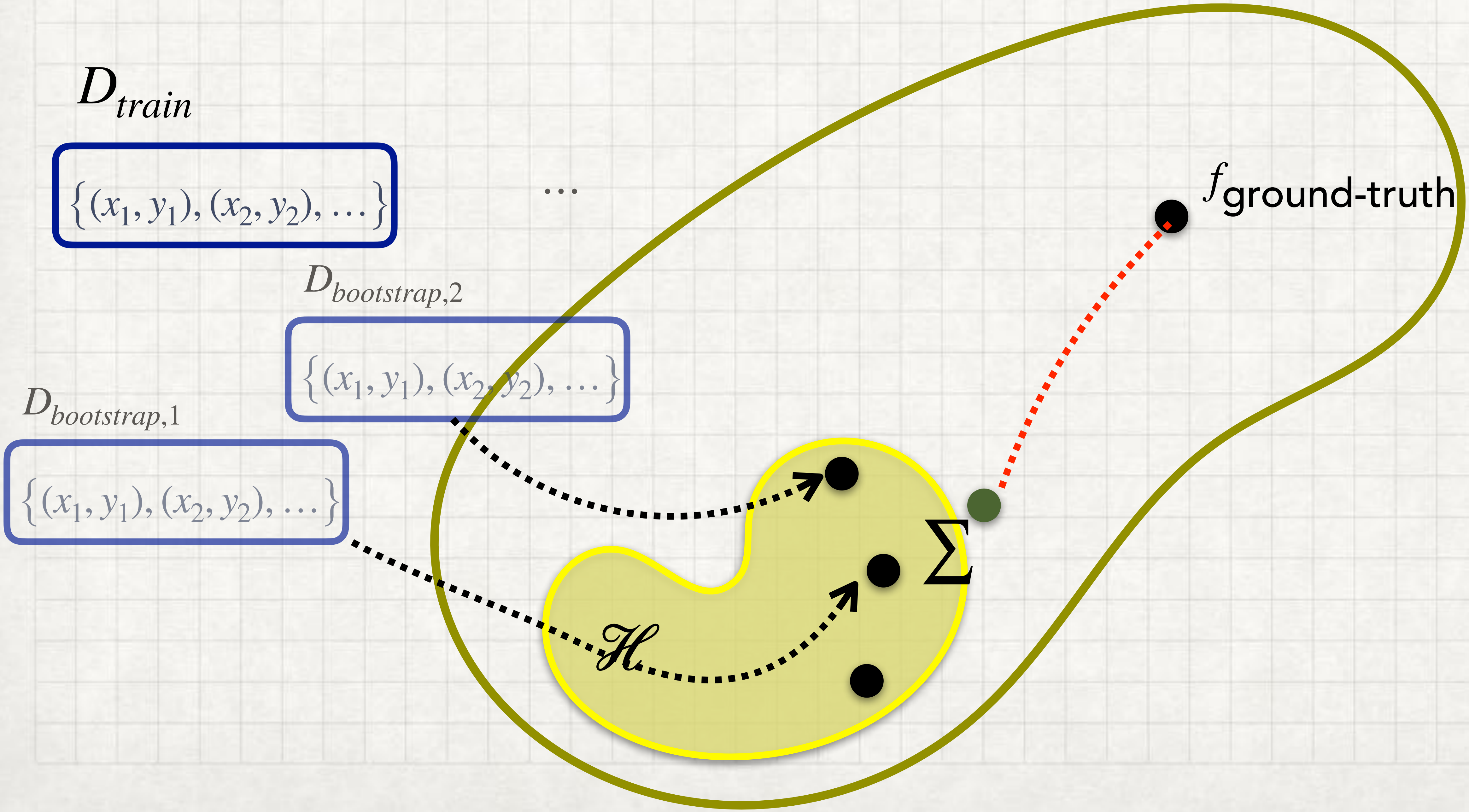
BAGGING IS AN ATTEMPT TO ACHIEVE "INFINITE" MODEL AVERAGING

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



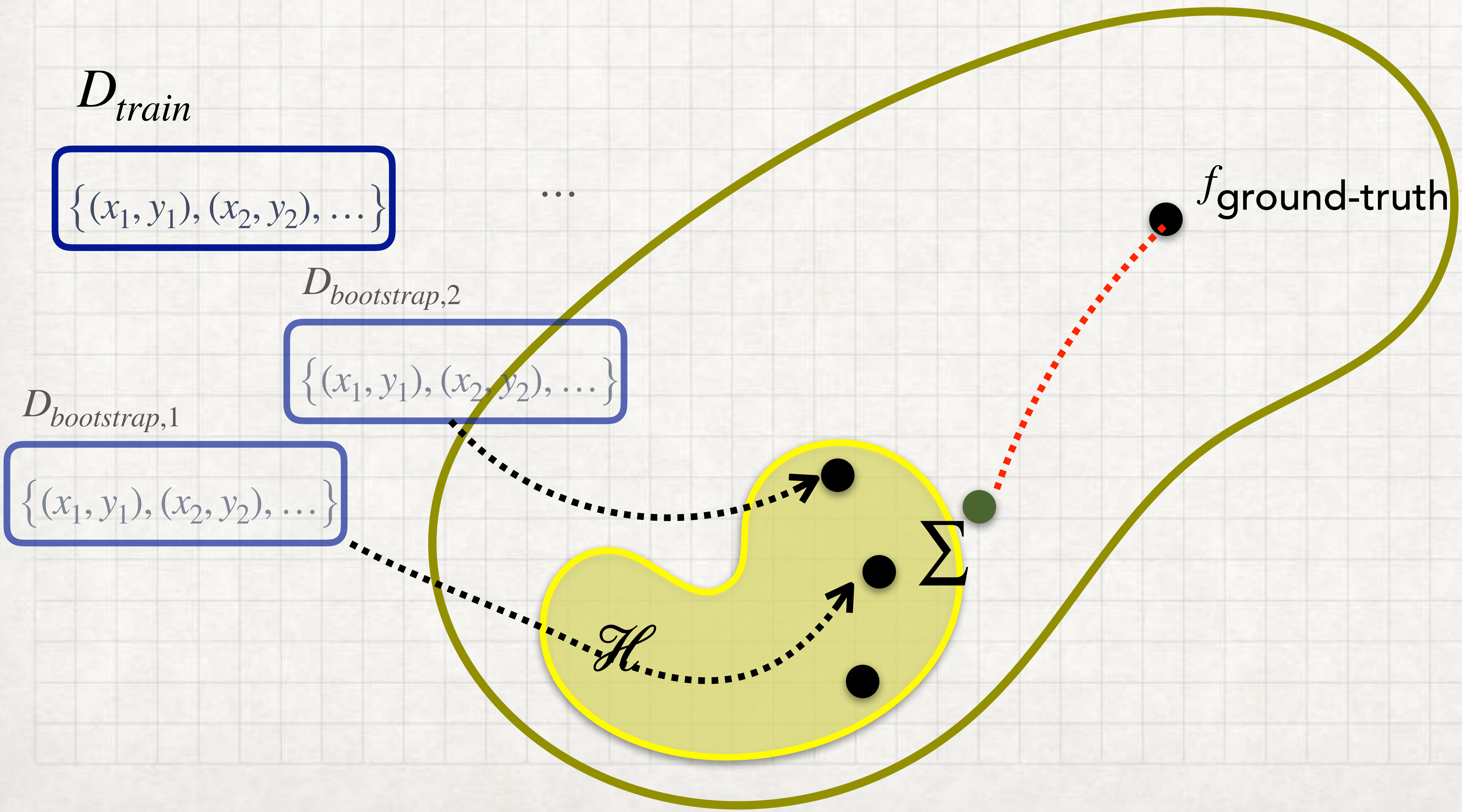
BOOTSTRAP SAMPLES

- 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.

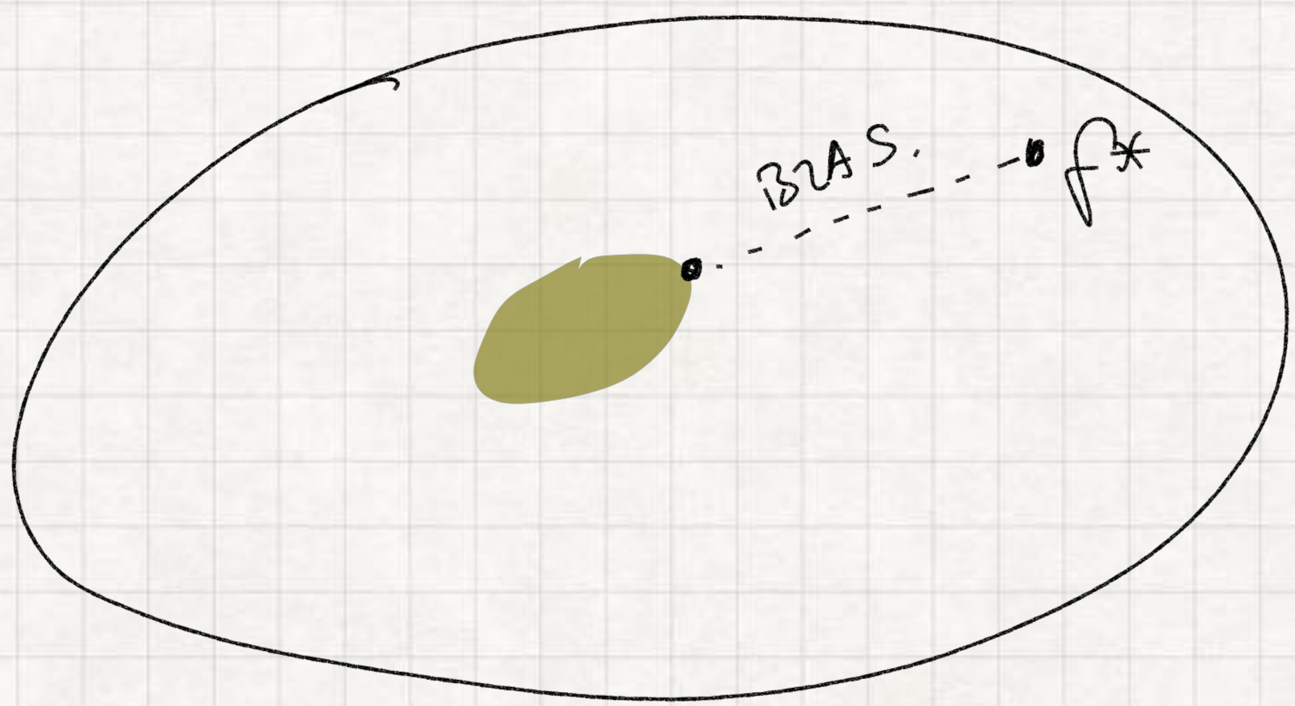
- 4.3 Observe Bias and Variance in sample model (if not done in the previous notebook study).

- 4.4 Bootstrap helps reduce variance in some settings.

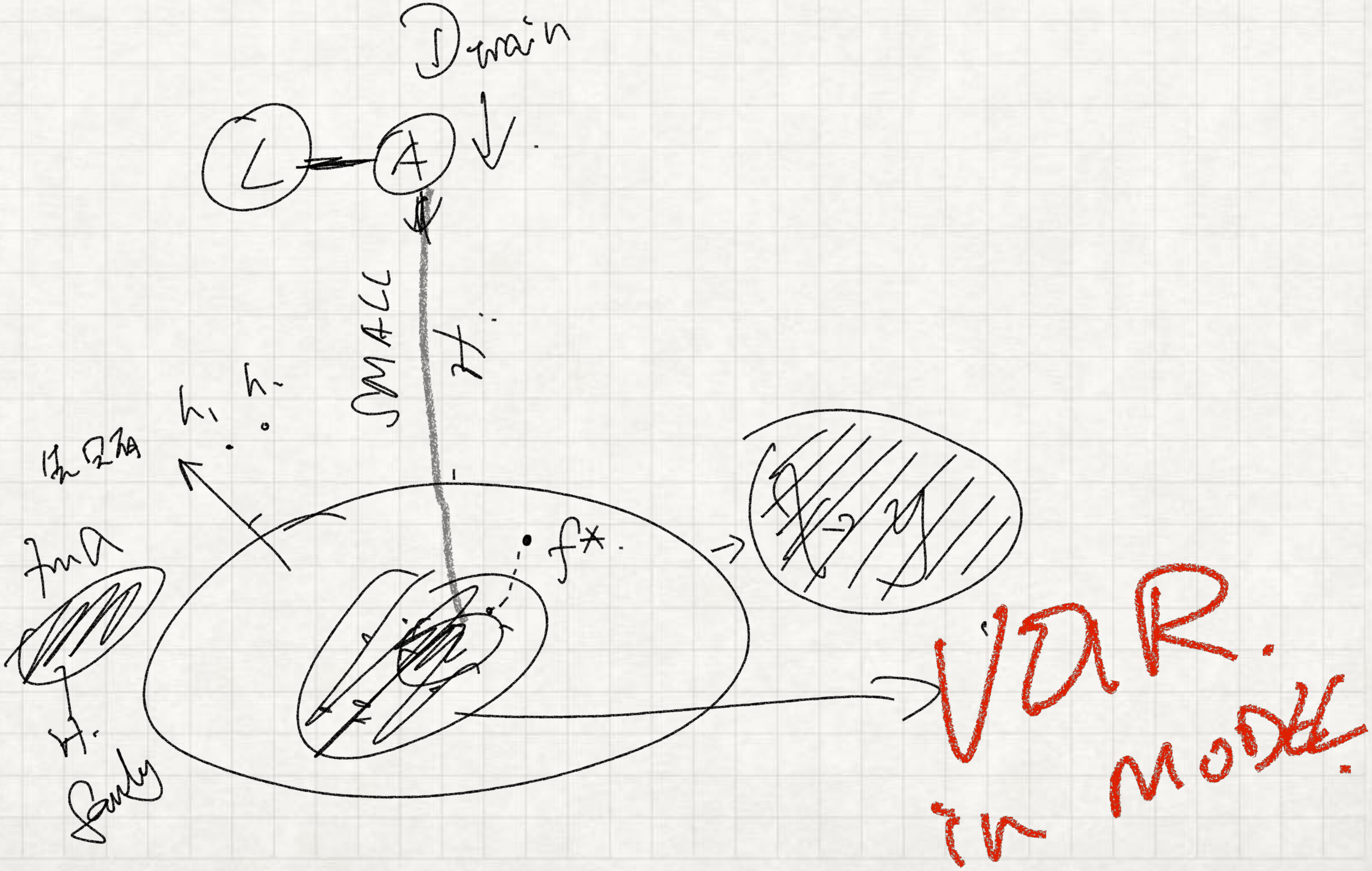
THANKS

WHAT WE HAVE LEFT?

- observation noise
- Training capability — partially why neural network works.



VARIANCE: BIG $\mathscr{H}$, BIG VARIANCE



NOTEBOOK TIME

- Compare bias/variance of two models