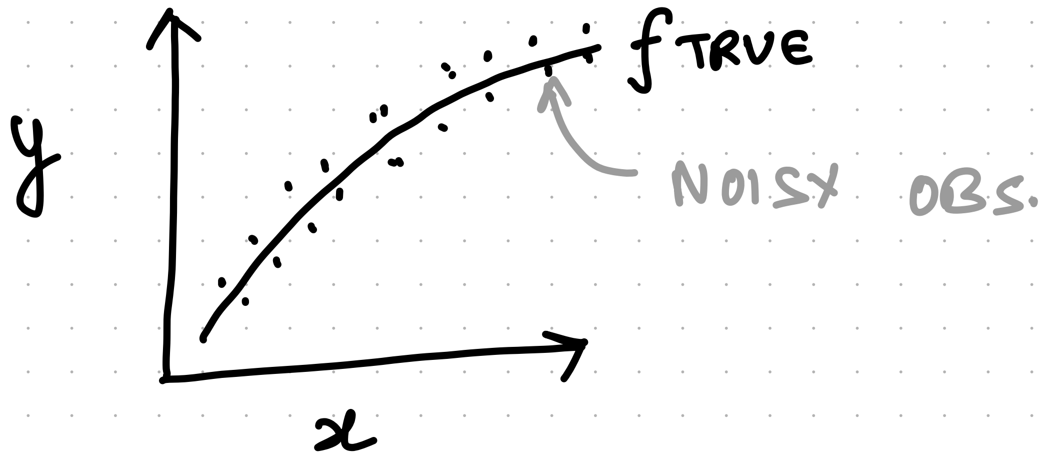
BIAS - VARIANCE

TRADE OFF

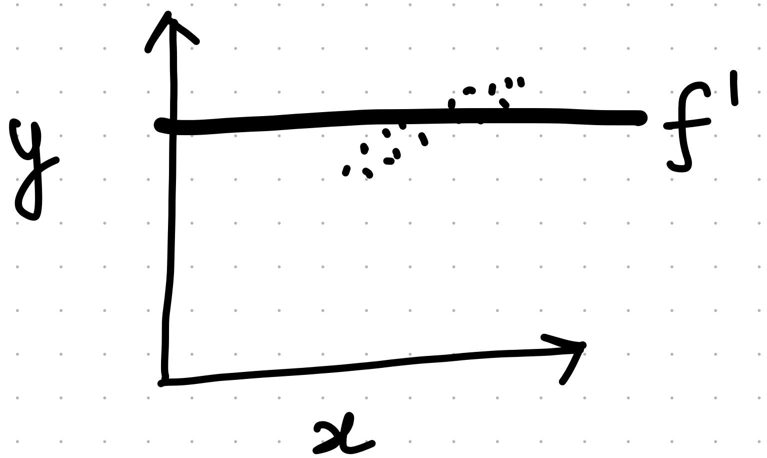
PART II

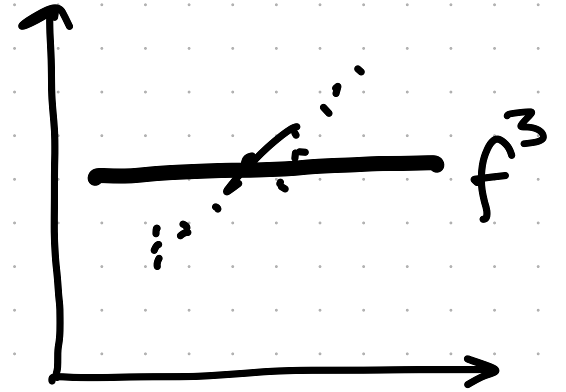


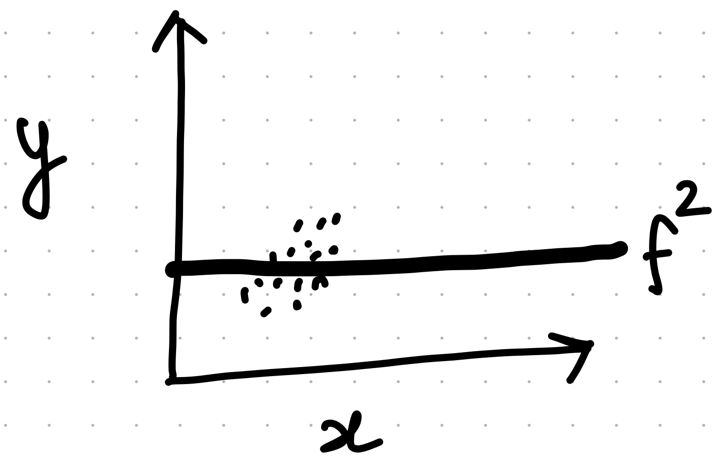
$$y_i = f_{TRUE}(x_i) + e_i$$

Observations = TRUE FUNCTION + Noise

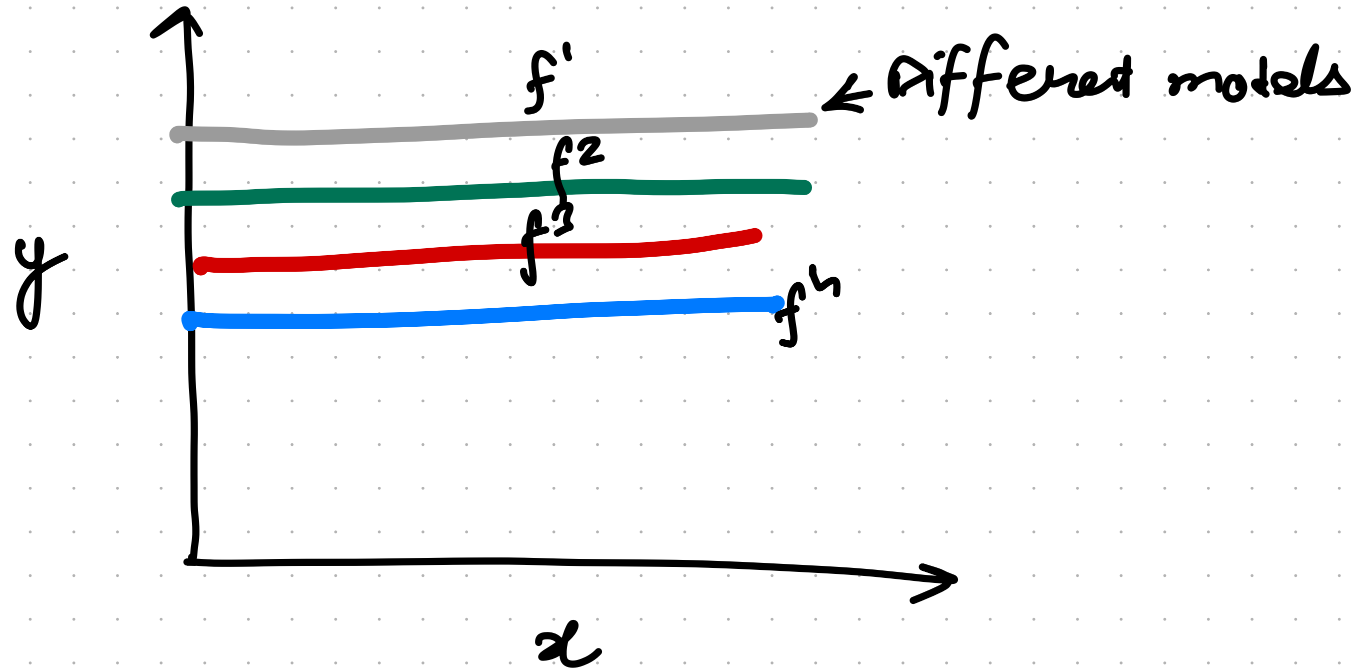
BIAS



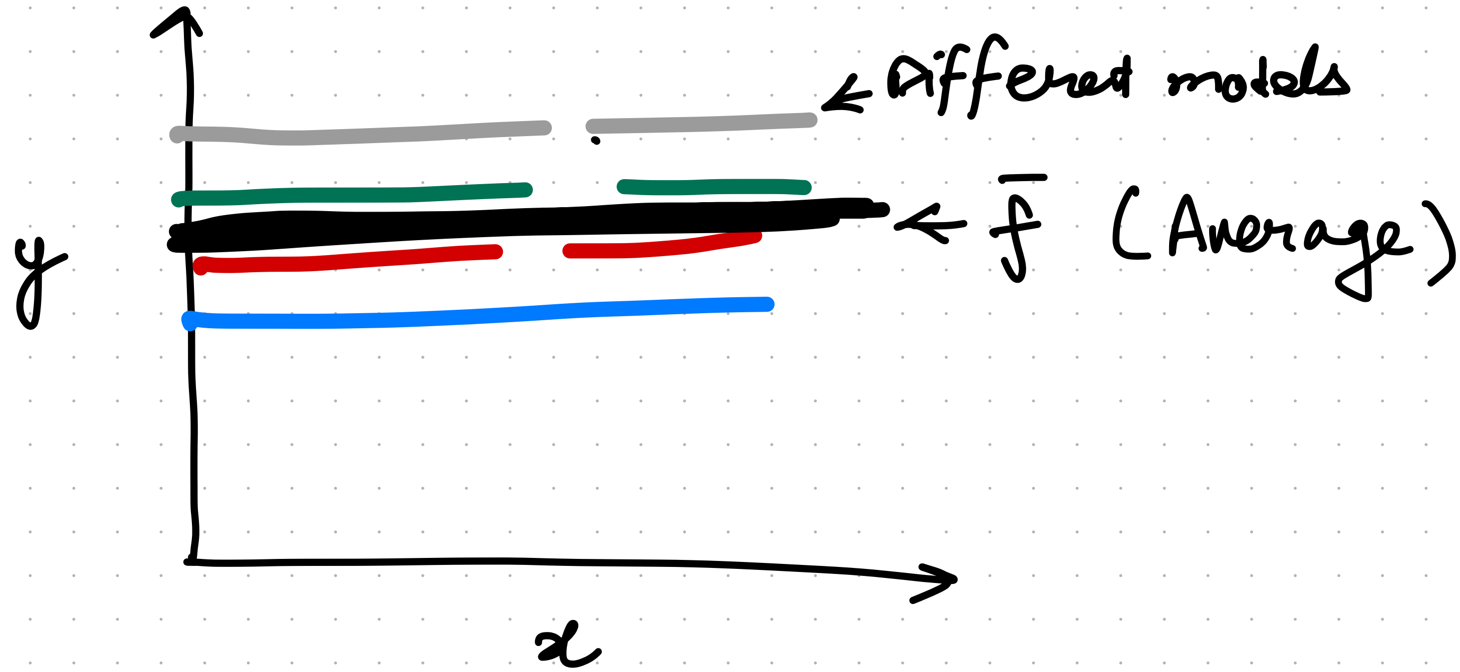




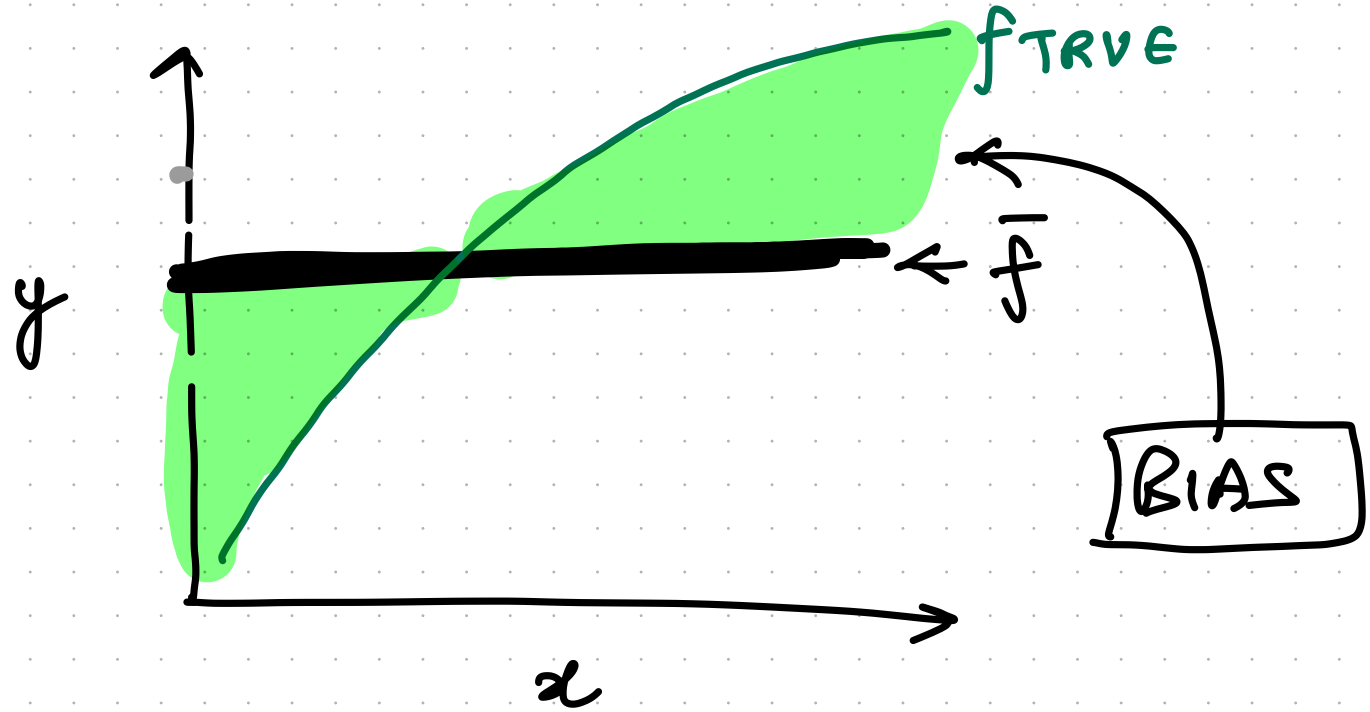
LEARN MODEL (say decision tree depth 1)
on different subsets of training data
(Assume we got a different subset from universe of possible training sets)



f^1
different models
f^2
f^3
f^4
y
x

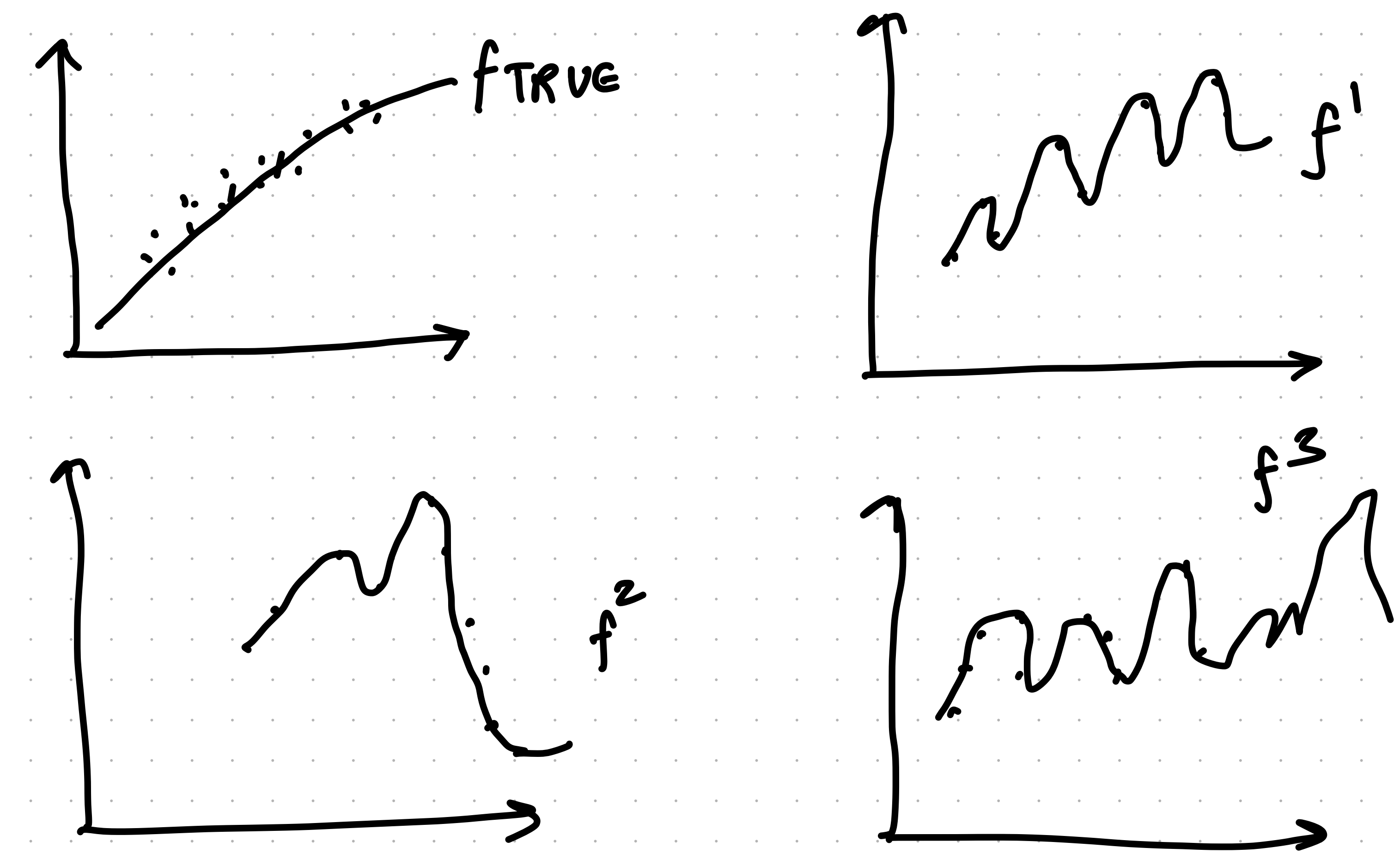


different models
$\bar{f}$ (Average)
y
x

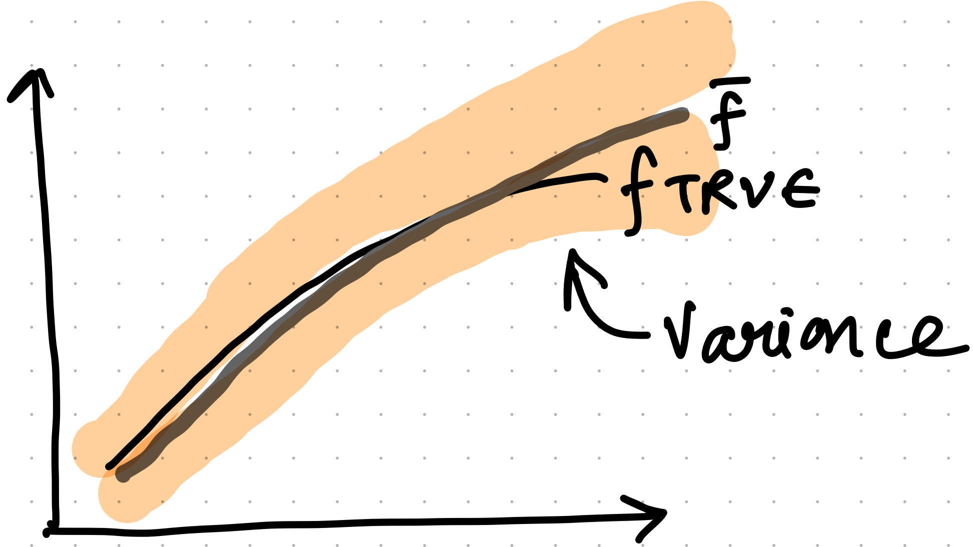


Bias = Deviation of our model (expectation over all possible training datsets) from true function

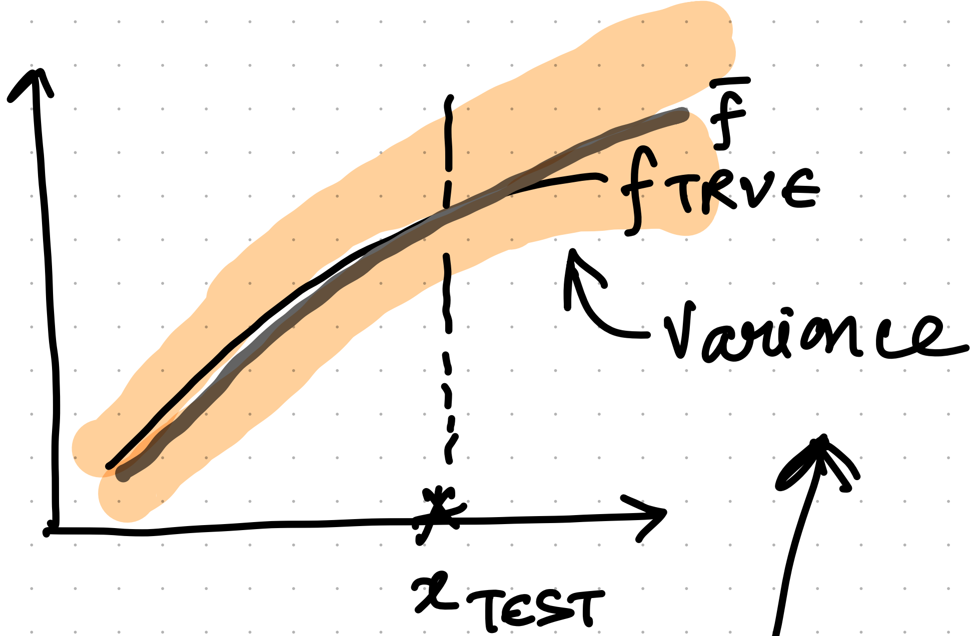
VARIANCE



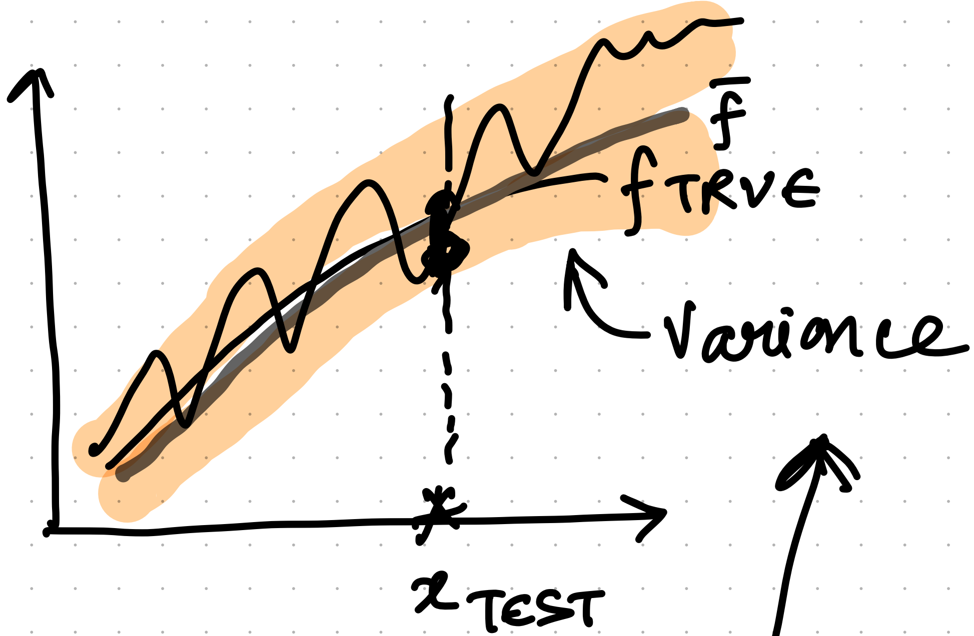
Same experiment as before (train model on different subsets)
<u>BUT</u> with sophistiated / complicated model (e.g. decision tree with ∞ depth)



Variance b/w different predictors



$$\left.\begin{aligned} f^1(x_{TEST}) &= 10 \\ f^2(x_{TEST}) &= 11 \\ f \cdots\ &= \end{aligned}\right\}\ \text{VARIANCE}$$



$f^1(x_{TEST}) = 10$

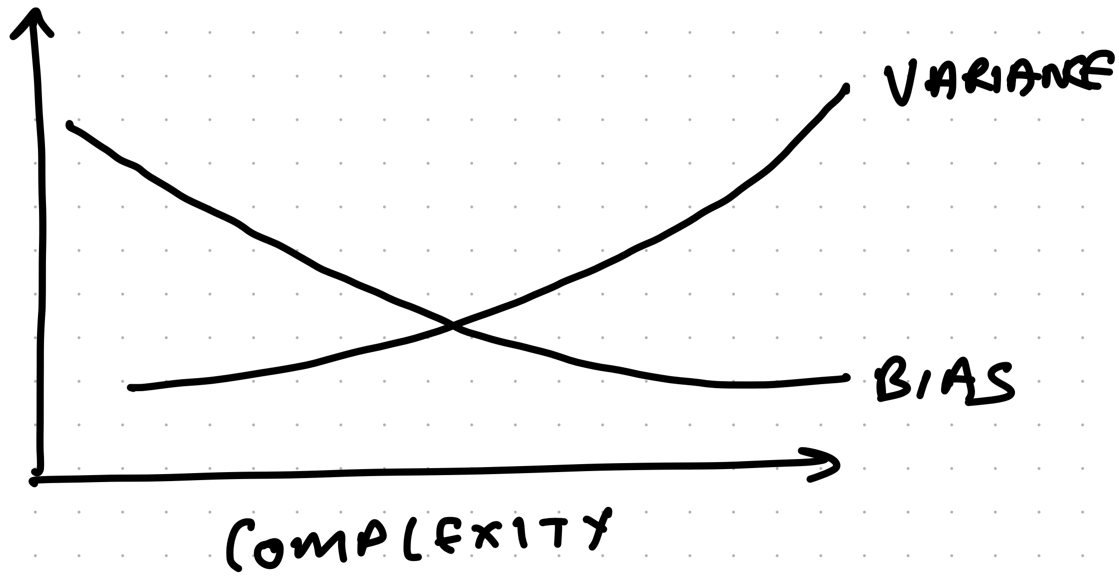
$f^2(x_{TEST}) = 11$

$f \cdot - - - =$

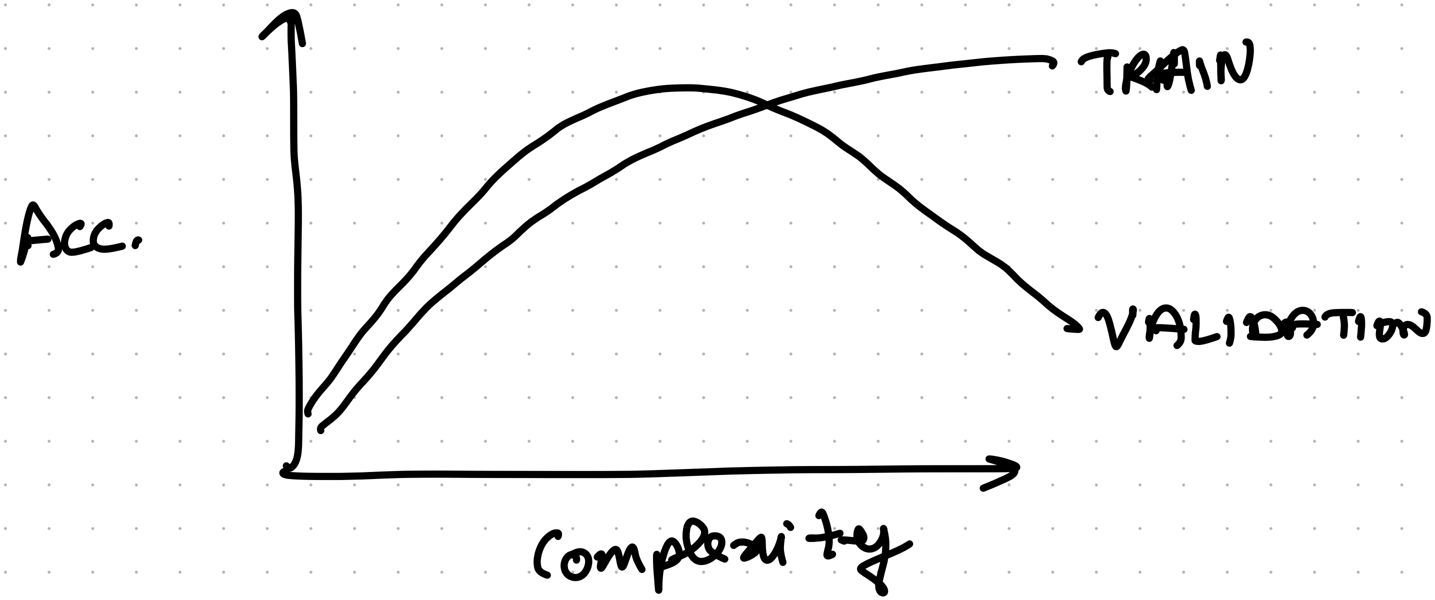
VARIANCE

3 SOURCES OF ERRORS

- OBSERVATION / IRREDUCIBLE
- BIAS
- VARIANCE



BIAS - VARIANCE
TRADE OFF



USING VALIDATION SET
TO FIND "RIGHT" COMPLEXITY

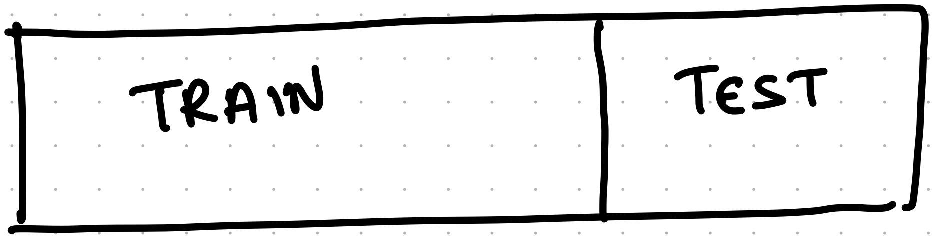
CROSS - VALIDATION

&

HYPERPARAMETER TUNING

STRATEGY #1

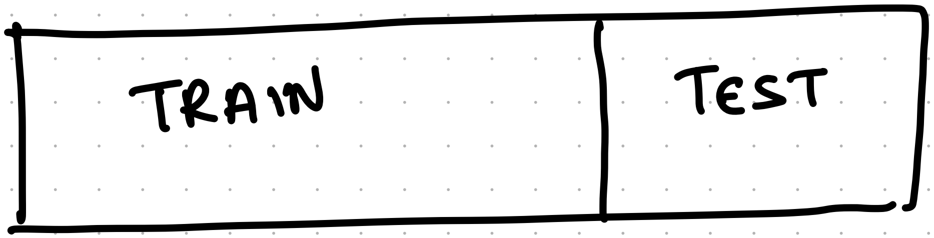
TRAIN-TEST SPLIT



- TRAIN ON TRAINING SET → MODEL
- EVALUATE MODEL ON TEST
- COMPUTE METRICS

STRATEGY #1

TRAIN - TEST SPLIT



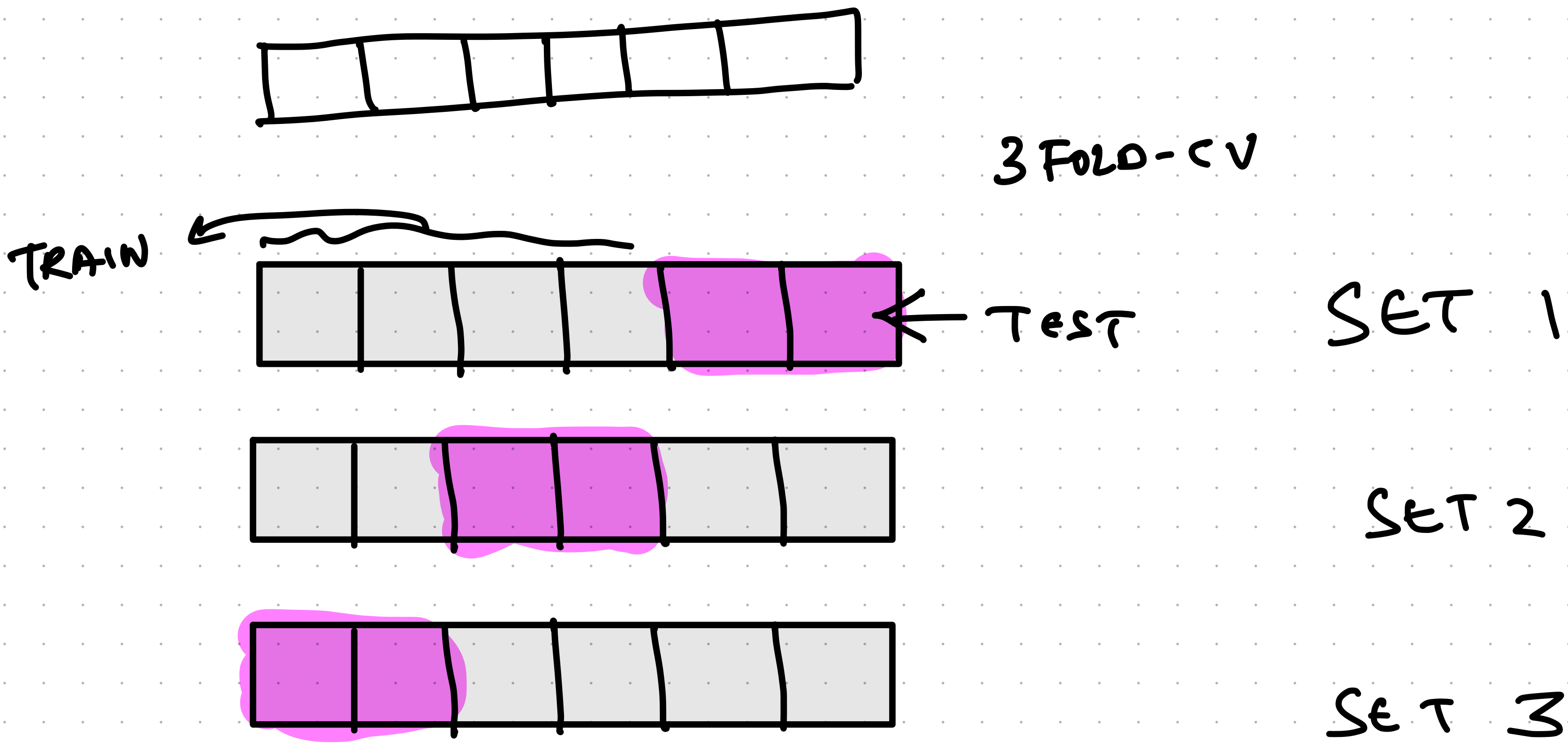
- TRAIN ON TRAINING SET → MODEL
- EVALUATE MODEL ON TEST
- COMPUTE METRICS

SHORTCOMINGS

- DOES NOT TEST ALL DATA POINTS
- NO WAY TO OPTIMIZE HYPERPARAMS.

STRATEGY #2

K-FOLD CV



FOR FOLD # IN #FOLDS

TRAIN ($\text{TRAINING}_{\text{FOLD \#}}$)

SCORE ($\text{TEST}_{\text{FOLD \#}}$)

REPORTING METRICS (e.g. Accuracy/ Error)

when using multiple models (e.g. K- FOLD CV)

TWO WAYS TO REPORT ACCURACY

(i) COMPUTE SCORE/METRIC FOR EACH FOLD

e.g. Accuracy fold 1 = 80% (2 samples)

fold 2 = 75% (... ...)

fold 3 = 85% (... ...)

Overall accuracy = 80% (MEAN)

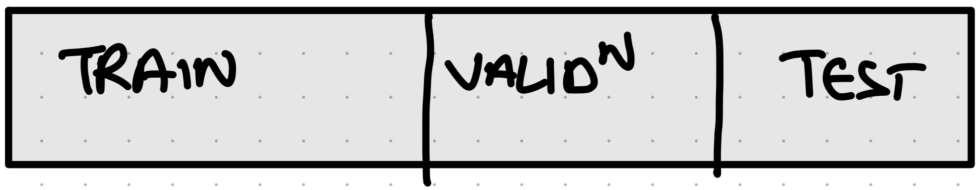
② CONCATENATE all predictions

FOLD 1	$\hat{y}_6$	y_6
	$\hat{y}_5$	y_5
FOLD 2	$\hat{y}_4$	y_4
	$\hat{y}_3$	y_3
FOLD 3	$\hat{y}_2$	y_2
	$\hat{y}_1$	y_1

$\Rightarrow$ Accuracy over 6 examples

STRATEGY #3

TRAIN – VALIDATION – TEST




```
MODELS = {}
SCORES = {}

FOR hp in hyperparams :
    MODELS [hp] = ALGO(** hp).fit (train)
    SCORES [hp] = ALGO.score (validation)
```

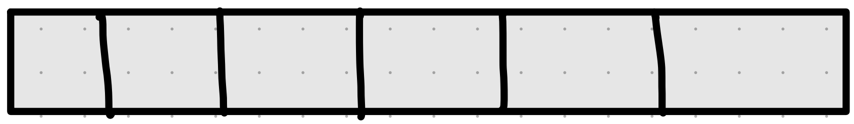
$$hp^* = \underset{hp}{\operatorname{argmax}}\ Scores$$

Depth	Valid^n
1	10
2	12
3	⋮
4	48
⋮	⋮
	10
100	8

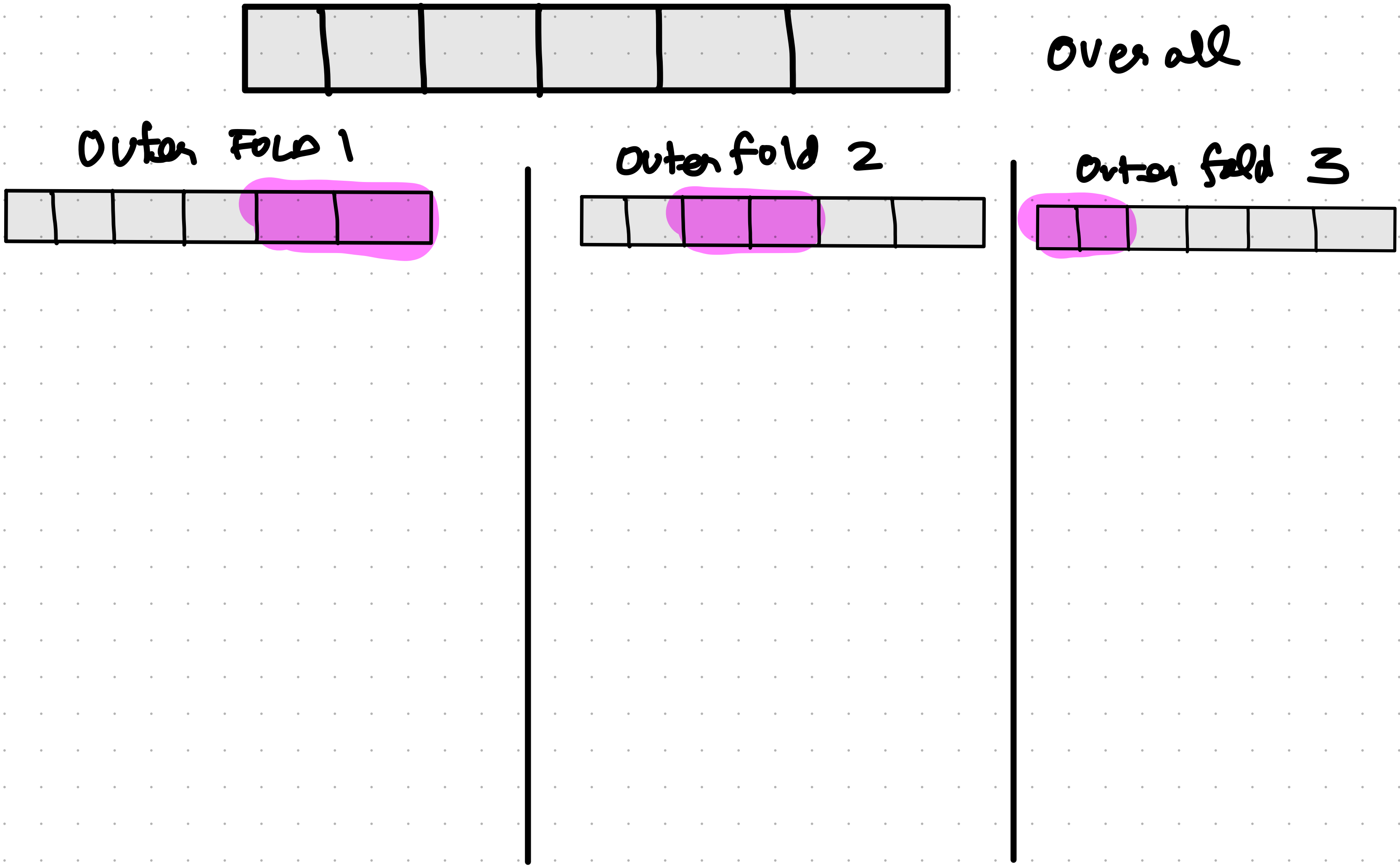
Use optimal HP & corresponding model to predict on test set.

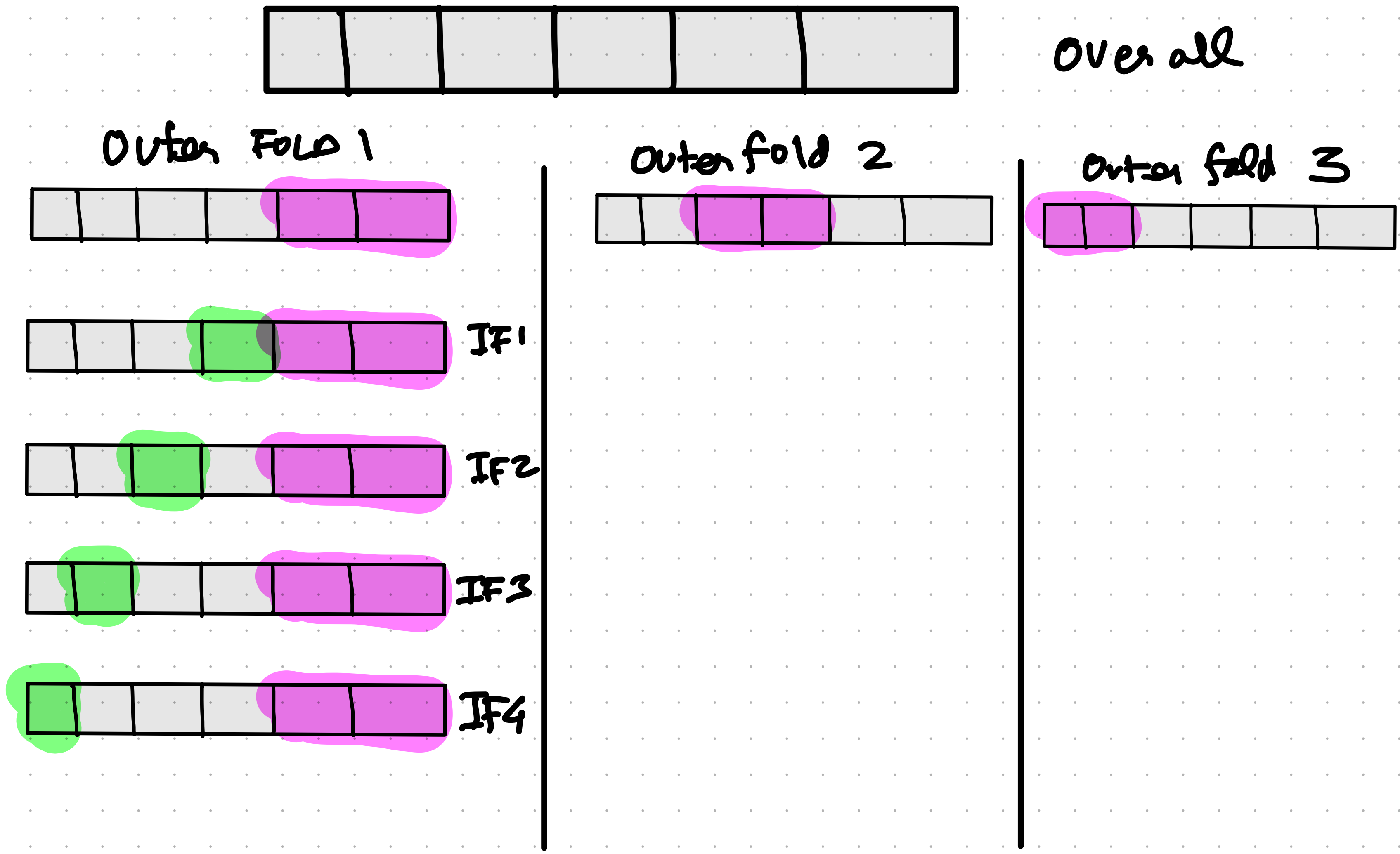
STRATEGY #4

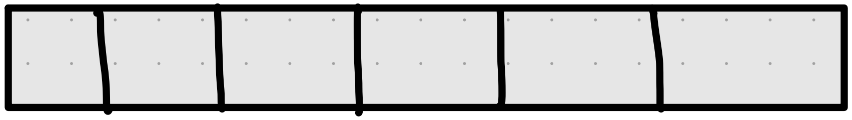
Nested Cross Validation



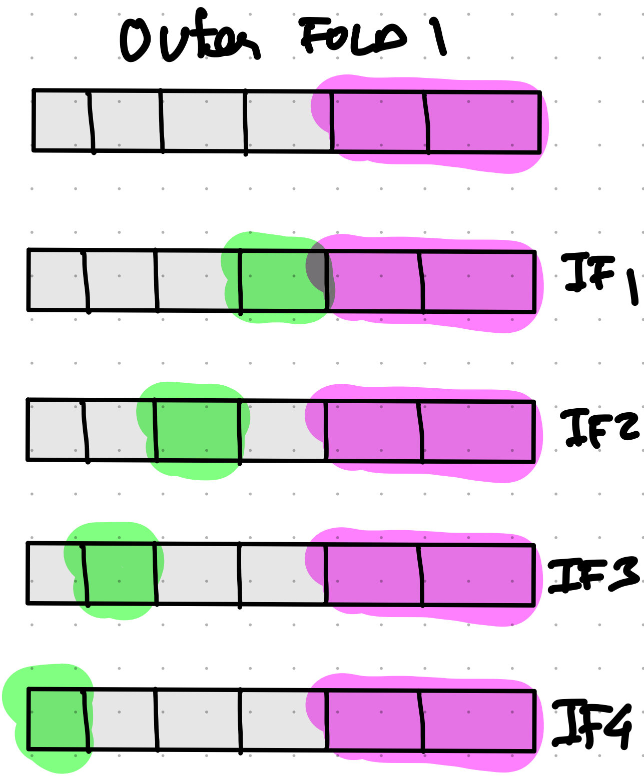
- Outer loop for data
- Inner loop for valid^{n} / hyperparameter tuning


Overall
Outer fold 1
Outer fold 2
Outer fold 3


Overall
Outer Fold 1
Outer fold 2
Outer fold 3
IF1
IF2
IF3
IF4



Overall

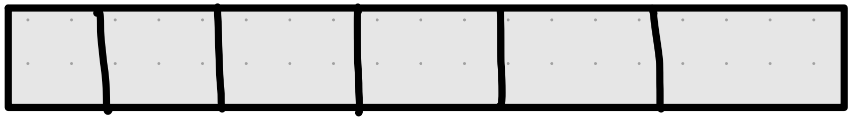


Assume single hp -depth

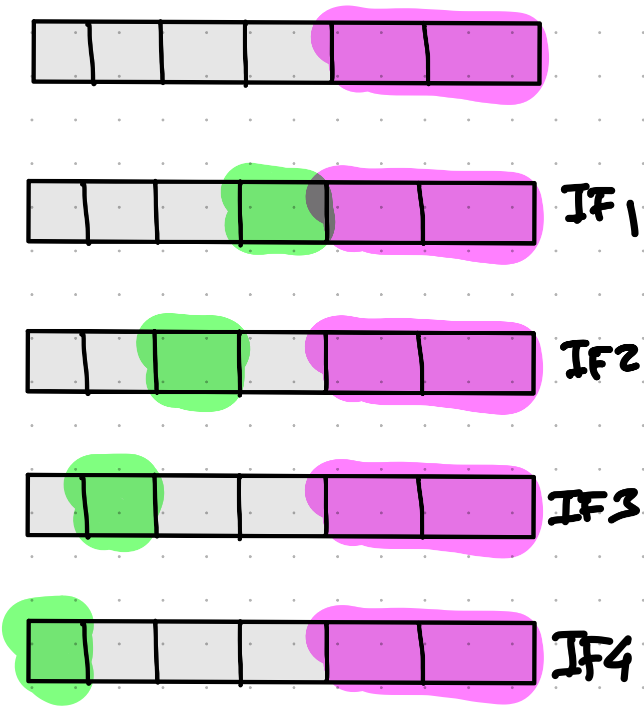
IF1		IF2		IF3		IF4	
d	s	d	s	d	s	d	s
1	80	1	70	1	80	1	70
2	90	2	90	2	100	2	100
3	60	3	70	3	60	3	70
4	30	4	80	4	30	4	80
5	60	5	60	5	20	5	40

d- depth

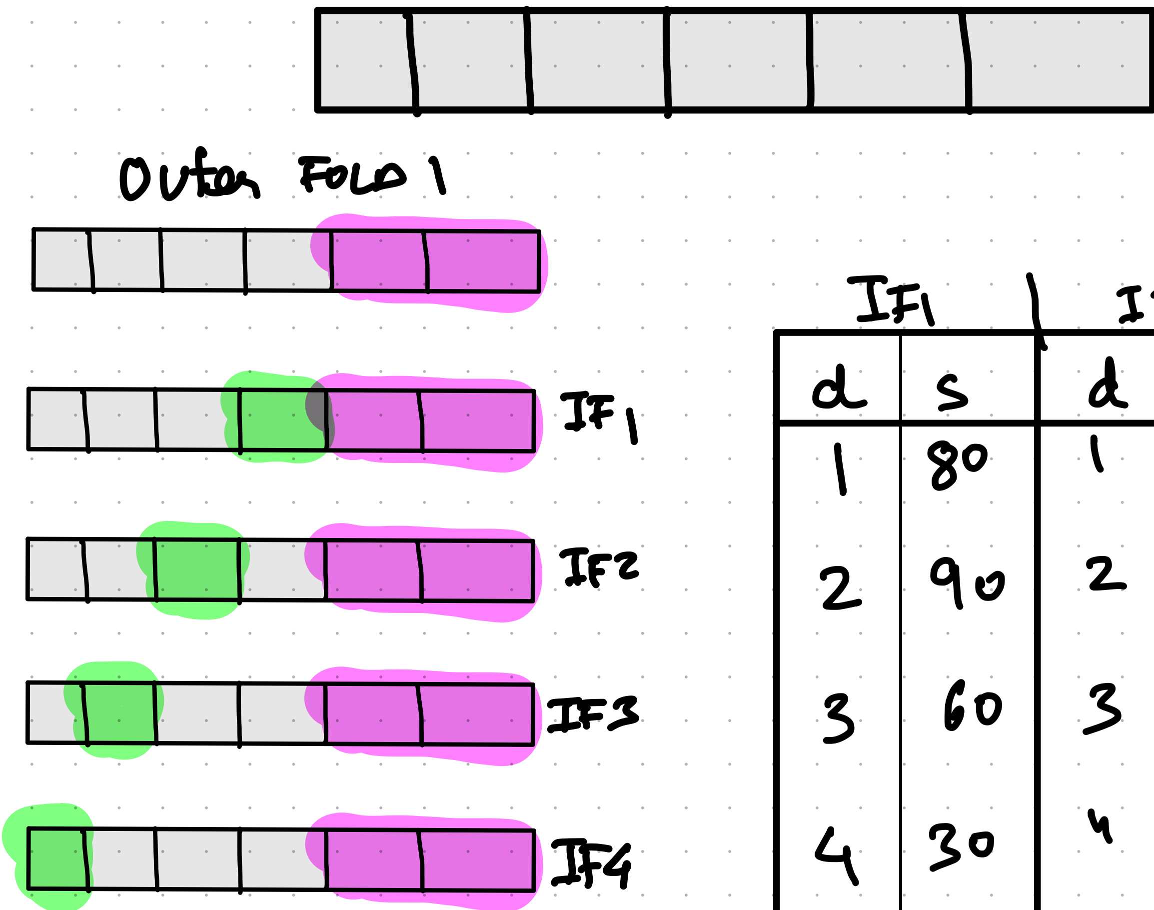
s - score on VALID^{N} SET



Outer Fold 1

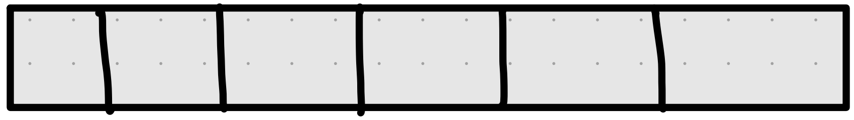


IF1		IF2		IF3		IF4		Avg	
d	s	d	s	d	s	d	s	d	s
1	80	1	70	1	80	1	70	1	75
2	90	2	90	2	100	2	100	2	95
3	60	3	70	3	60	3	70	3	65
4	30	4	80	4	30	4	80	4	55
5	60	5	60	5	20	5	40	5	45

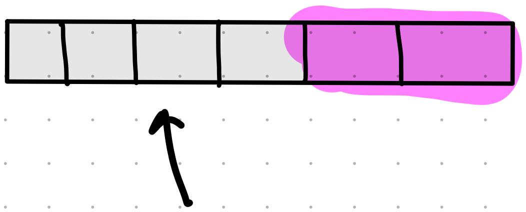


IF1		IF2		IF3		IF4		Avg	
d	s	d	s	d	s	d	s	d	s
1	80	1	70	1	80	1	70	1	75
2	90	2	90	2	100	2	100	2	95
3	60	3	70	3	60	3	70	3	65
4	30	4	80	4	30	4	80	4	55
5	60	5	60	5	20	5	40	5	45

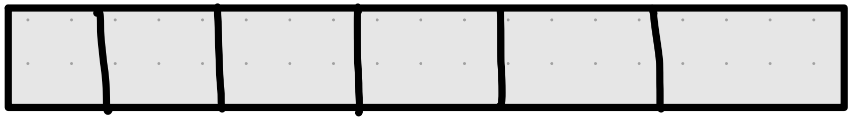
$$depth^* = 2$$



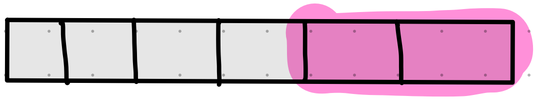
Outer Fold 1



TRAIN with $depth^* = 2$

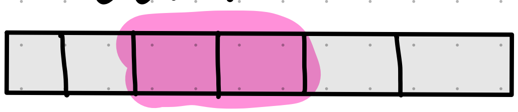


Outer FOLD 1



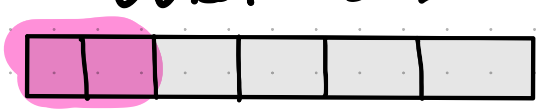
TRAIN with $depth^* = 2$

Outer FOLD 2



TRAIN with depth optimized on 4 inner folds for outer fold 2

Outer FOLD 3



TRAIN with depth optimized on 4 inner folds for outer fold 3