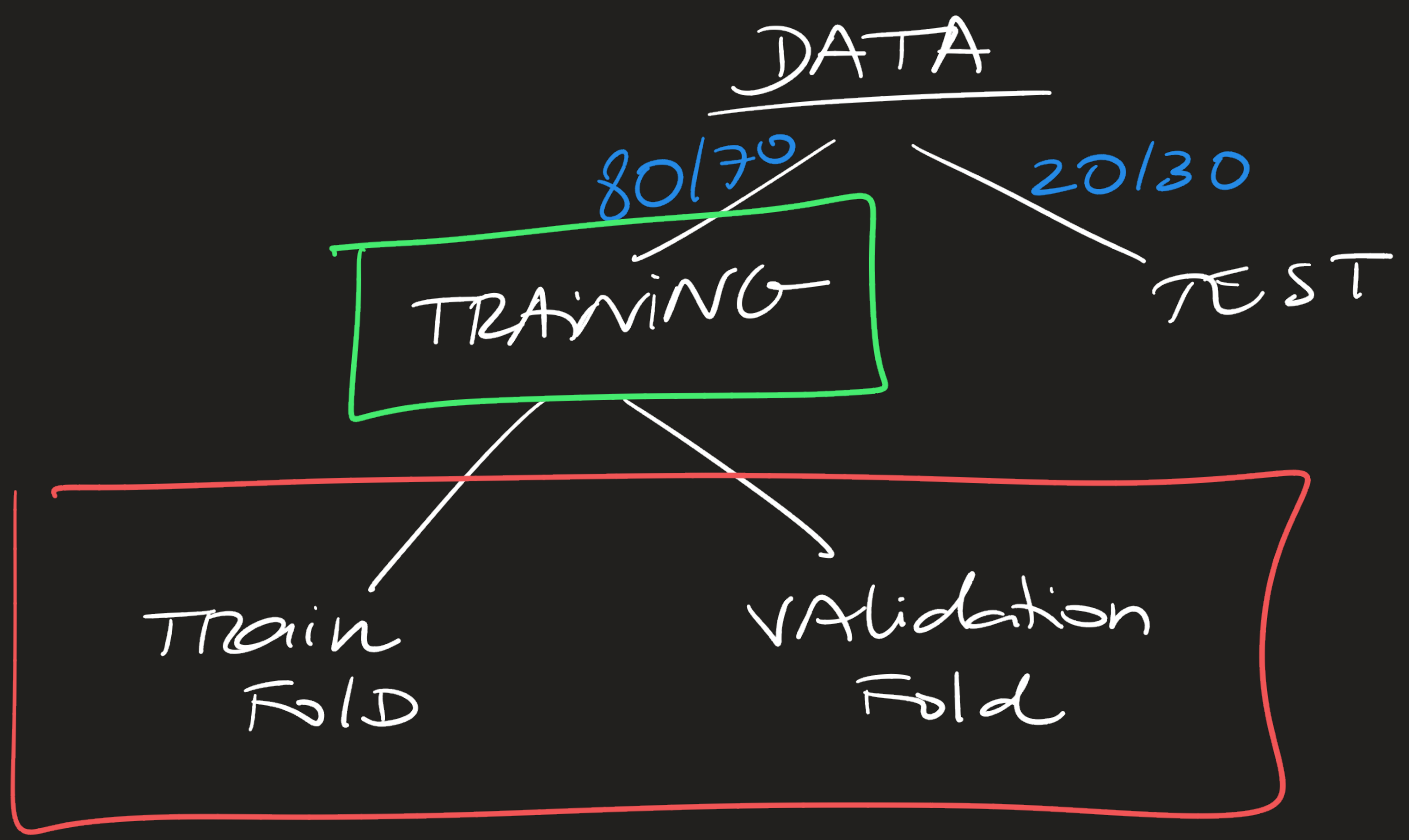
Cross-validation

It is used to choose the hyperparameter values using a validation set.



CV.

1) K-FOLD CROSS-VALIDATION

TRAINING DATA (80%)

S_1	S_2	S_3	S_4

$K=4$

K value

5–20

a) Shuffle data and then partition it into K equally-sized folds.

polynomial model order

for M in {2, 3, 4, ..., 100}:

for λ in [0, 10]:

{M value, λ value} = hyperparameter set i

1) Train set = $\{S_1, S_2, S_3\} \rightarrow W_1$
val. set = $\{S_4\}$
MSE-train = 1, MSE-VAL = 10

2) Train set = $\{S_1, S_3, S_4\} \rightarrow W_2$
val. set = $\{S_2\}$
MSE-train = 100, MSE-VAL = 102

3) Train set = $\{S_2, S_3, S_4\} \rightarrow W_3$
val. set = $\{S_1\}$
MSE-train = 0.01, MSE-VAL = 5

4) Train set = $\{S_1, S_4, S_2\} \rightarrow W_4$
val. set = $\{S_3\}$
MSE-train = 5, MSE-VAL = 50

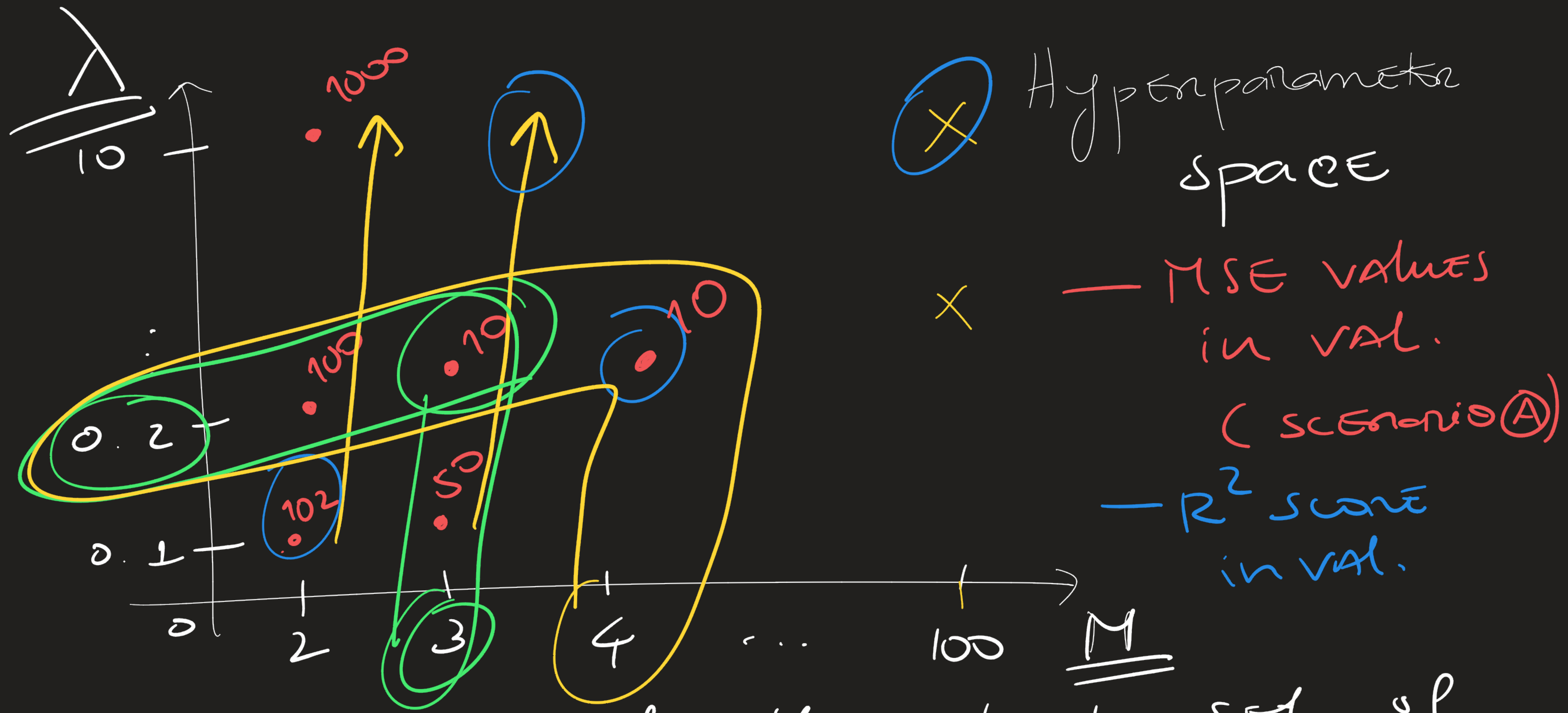
For each configuration of [M, λ], we report a performance measure in VALIDATION.

A) select the worst performance score in val. set

SCORE-VAL = 102

B) select the average performance in val. set

$$\text{SCORE-VAL} = \frac{10 + 102 + 5 + 50}{4}$$



Once we identify the best set of hyperparameter values, we train the final model using the entire training set (80%) w/ the best hyperparameters values.

Experimental Design

1) Factorial or Grid search design

↳ tries every combination.

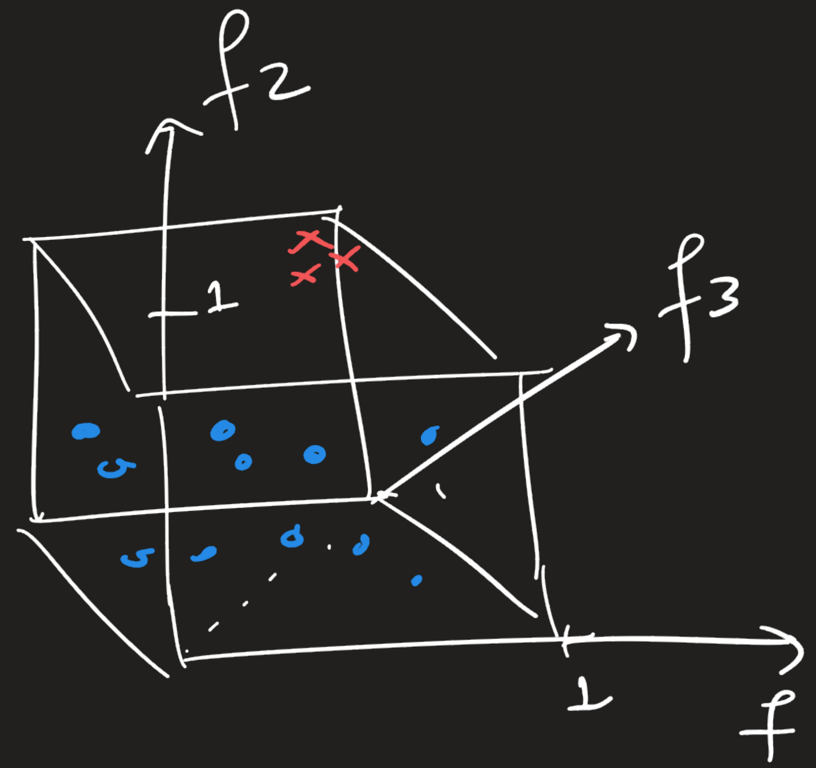
$\boxed{\text{2) Random Grid Search}}$ MOST popular

↳ selects a subset of configurations $\{M, \lambda\}$ at random.

Curse of Dimensionality

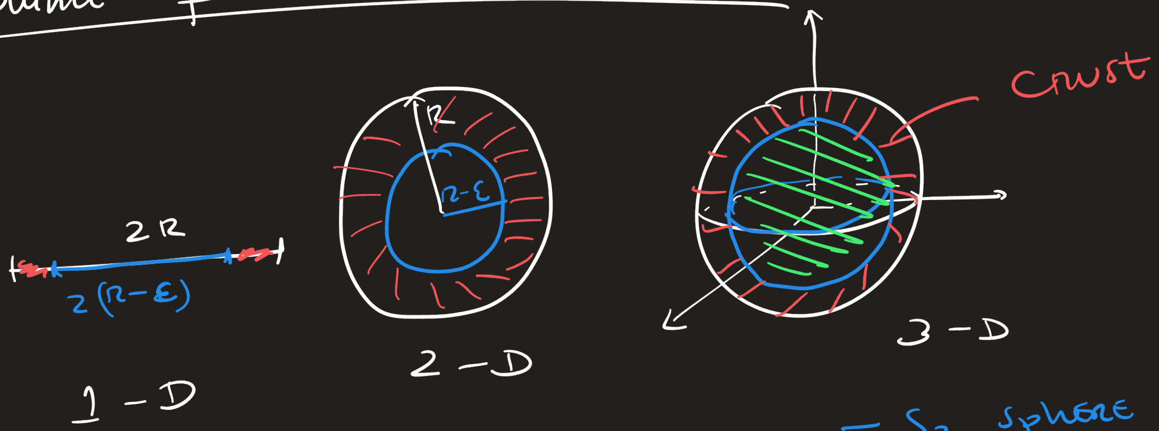
Features ≡ dimensions

in your feature space.



∘ training samples
× test samples

1) Volume of the crust example



— S_2 sphere (Radius $R-\varepsilon$)

— S_1 sphere (Radius R)

$$\boxed{\varepsilon << R}$$

$$\frac{V_{crust}}{V_{S_1}} \xrightarrow[\text{dimensions} \to \infty]{}$$

V_{sphere} w/ radius R in D-dimensions $= \dfrac{R^D \cdot \pi^{D/2}}{\Gamma(D/2+1)}$

$$\Gamma(x) = (x-1)!$$

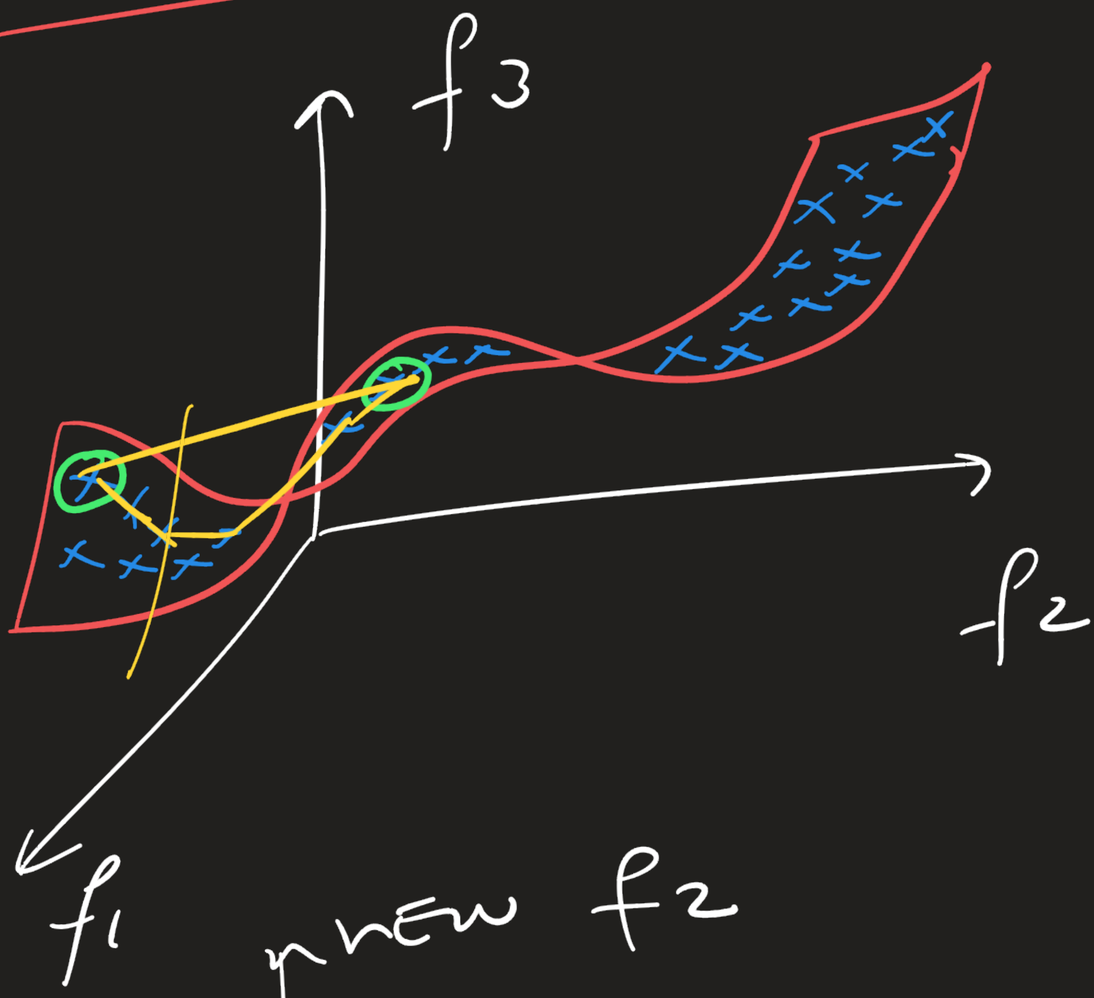
$$\frac{V_{crust}}{V_{S_1}} = \frac{V_{S_1} - V_{S_2}}{V_{S_1}}$$

$$= 1 - \frac{V_{S_2}}{V_{S_1}}$$

$$= 1 - \frac{\dfrac{(R-\varepsilon)^D \cdot \cancel{\pi^{D/2}}}{\cancel{\Gamma(D/2+1)}}}{\dfrac{R^D \cdot \cancel{\pi^{D/2}}}{\cancel{\Gamma(D/2+1)}}}$$

$$= 1 - \frac{(R-\varepsilon)^D}{R^D} = 1 - \left(1 - \frac{\varepsilon}{R}\right)^D \xrightarrow[D \to \infty]{} \underline{\underline{1}}$$

Dimensionality Reduction



input space

× – training samples

: manifold (unknown)

but learnable

manifold learning.