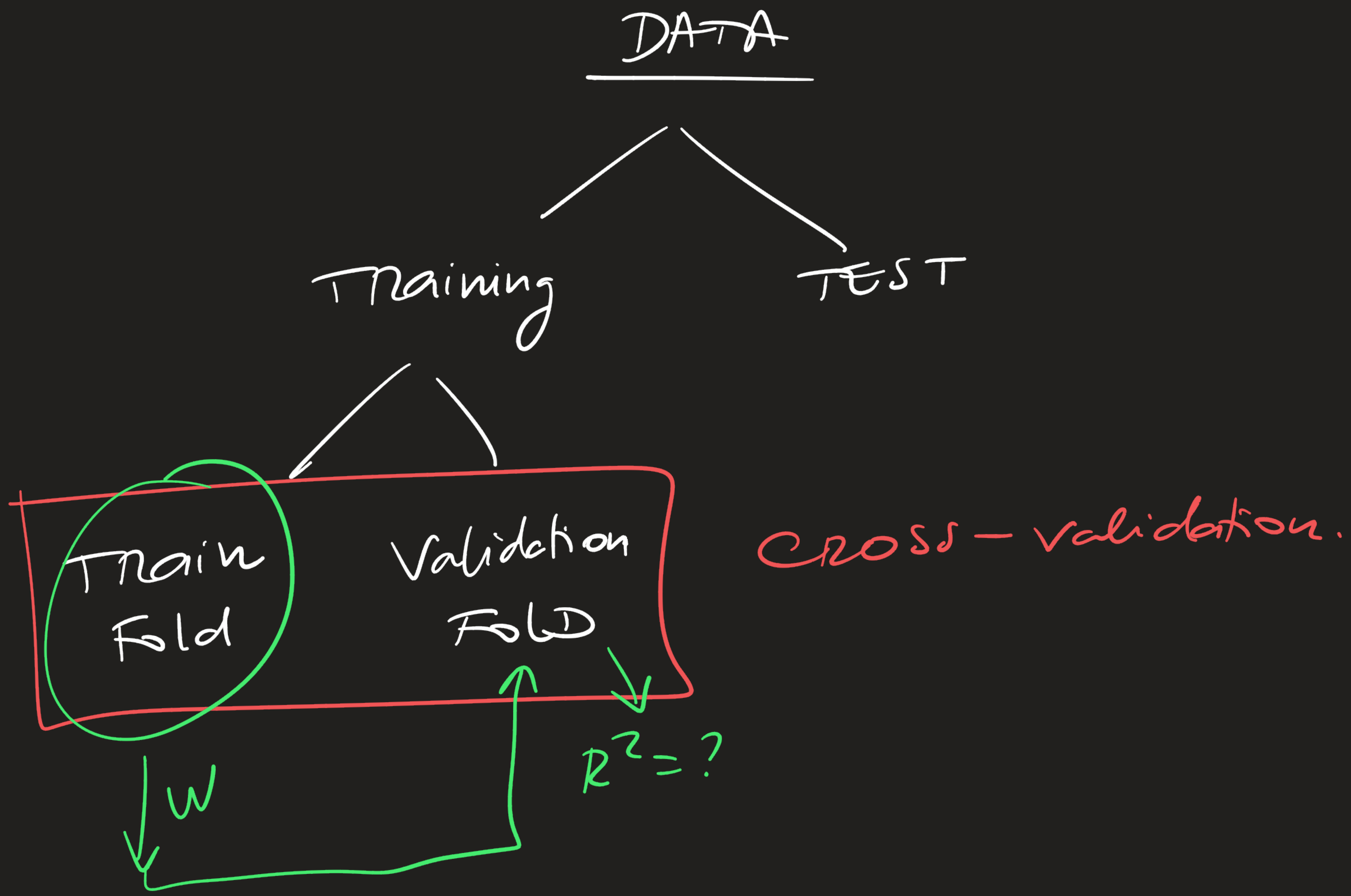
Ways to Mitigate Overfitting

1. Regularization
   ↳ adds a penalty term on the parameters, w, in the objective function.

2. Add more data!

3. Apply Occam's Razor Principle
   ↳ select the simplest high-performing model.

4. Cross-validation.
   ↳ Makes use of data to help select hyperparameters.



DATA
Training
TEST
Train Fold
Validation Fold
cross-validation.
w
$R^2 = ?$

Regularization

① Ridge Regularizer

$$R_2 = \|w\|_2 = \left(w_0^2 + w_1^2 + \dots + w_n^2 \right)^{1/2}$$

② Lasso Regularizer → promotes sparsity.

↓ drives some of the coefficients to exactly zero.

$$R_1 = \|w\|_1 = |w_0| + |w_1| + \dots + |w_n|$$

③ Elastic Net Regularizer

$$R_{EN} = \alpha \cdot R_2 + (1-\alpha) \cdot R_1 \quad , \alpha \equiv \text{hyperparameter}$$

$$\alpha \in [0, 1]$$

Regularized objective function

$$\boxed{J(w) = \|t - Xw\|_2^2 + \lambda \cdot R_2^2(w)}$$

$$\boxed{= \|t - Xw\|_2^2 + \lambda \cdot \|w\|_2^2}$$

Ridge MSE OR Ridge Regression objective function.

Regularizer hyperparameter

Solution for w :

$$\frac{\partial J}{\partial w} = 0 \iff \boxed{w = \left(X^T X + \lambda \cdot I\right)^{-1} \cdot X^T \cdot t}$$

w is $(M+1) \times 1$

X is $N \times (M+1)$

I is $(M+1) \times (M+1)$

λ is scalar

t is $N \times 1$

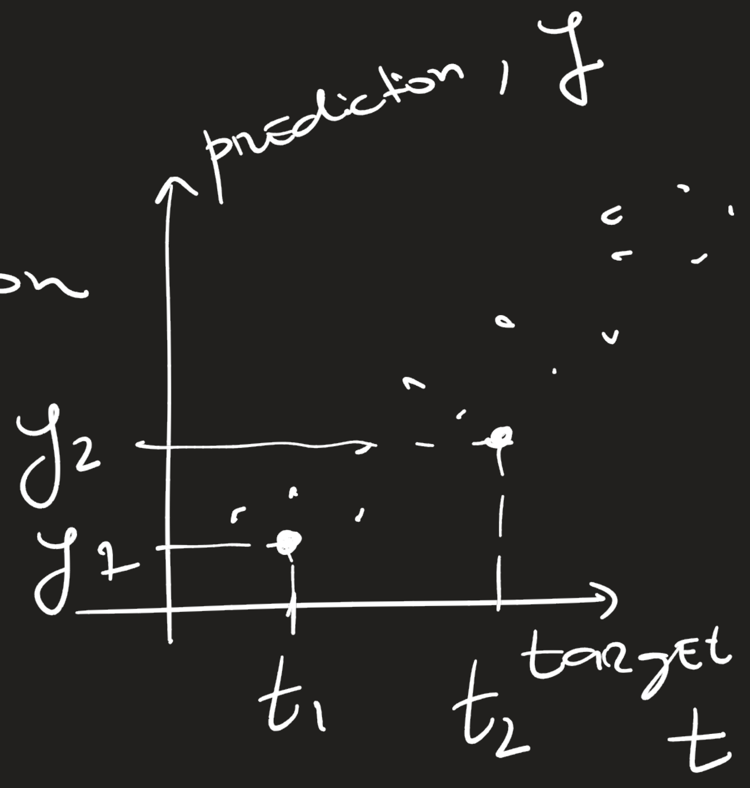
$$y = w_0 + w_1 \cdot x^1 + w_2 x^2 + \dots + w_M \cdot x^M$$

$$= w_0 + \underbrace{\cancel{w_1}}_{=0} \cdot f_1 + w_2 \cdot f_2 + \dots + \underbrace{\cancel{w_M}}_{=0} \cdot f_M$$

Performance Metrics for Regression Tasks

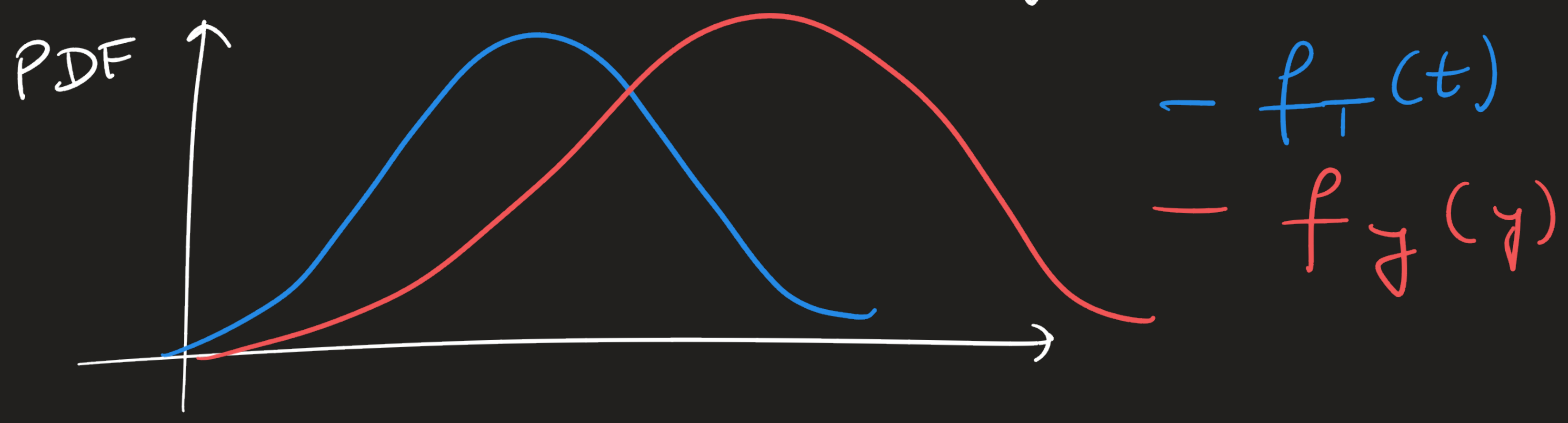
→ Error - based metrics

→ Coefficient of determination
$R^2 \in [0, 1]$



→ R^2 of the Q-Q plot
(quantile - quantile plot).

we look at t and y as R.V.



[0.1, 0.7, 0.9]

quantiles of y
(sorted values of y)

· — ideal Q-Q plot

R^2 of Q-Q is more optimistic.

[0.5, 0.9, 1]

quantiles of t

(sorted values of t)

$t = [1, 0.5, 0.9]$

$y = [0.9, 0.7, 0.1]$