input : $x_i \leftarrow$ time

target : $t_i \leftarrow$ foam height

Polynomial Model $M = 3$

$$y = w_0 + w_1 x + w_2 x^2 + w_3 x^3$$

solution : $w = (X^T X)^{-1} \cdot X^T \cdot t$

Exponential Model

$$y(x) = e^{w_0 + w_1 . x} \quad , \quad x \equiv \text{time}$$

$$\Leftrightarrow \quad \ln(y) = w_0 + w_1 . x$$

$$w = (X^T X)^{-1} X^T \ln(t)$$

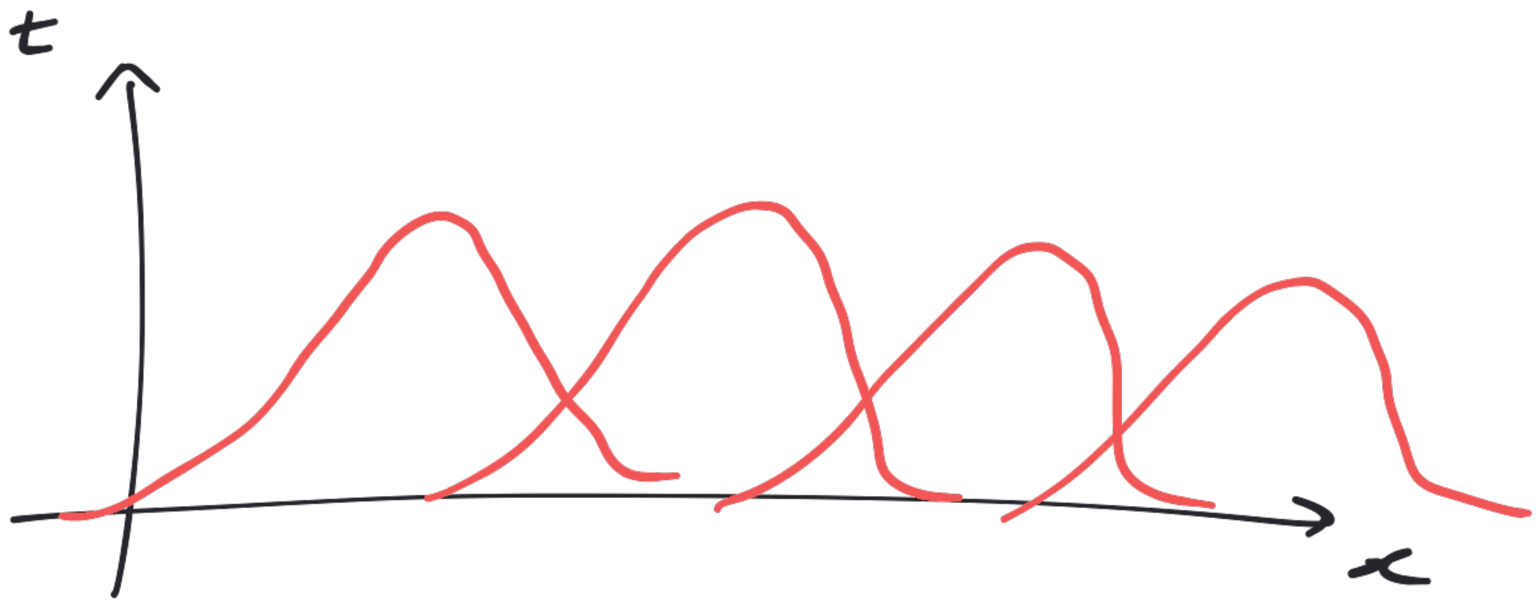
P2

$$y = w_0 + w_1 \phi_1(x) + w_2 \phi_2(x) + w_3 \phi_3(x) + w_4 \phi_4(x)$$

$$\phi_i(x) = e^{-\frac{(x-\mu_i)^2}{2\sigma^2}} \quad , \quad \sigma = 0.01$$

$$X = \begin{bmatrix} 1 & e^{-\frac{(x_1-\mu_1)^2}{\sigma^2}} & \cdots & e^{-\frac{(x_1-\mu_4)^2}{\sigma^2}} \\ 1 & e^{-\frac{(x_2-\mu_1)^2}{\sigma^2}} & \cdots & e^{-\frac{(x_2-\mu_4)^2}{\sigma^2}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\frac{(x_N-\mu_1)^2}{\sigma^2}} & \cdots & e^{-\frac{(x_N-\mu_4)^2}{\sigma^2}} \end{bmatrix}$$

$N \times 5$



Ways to Avoid Overfitting

1. Add more data!

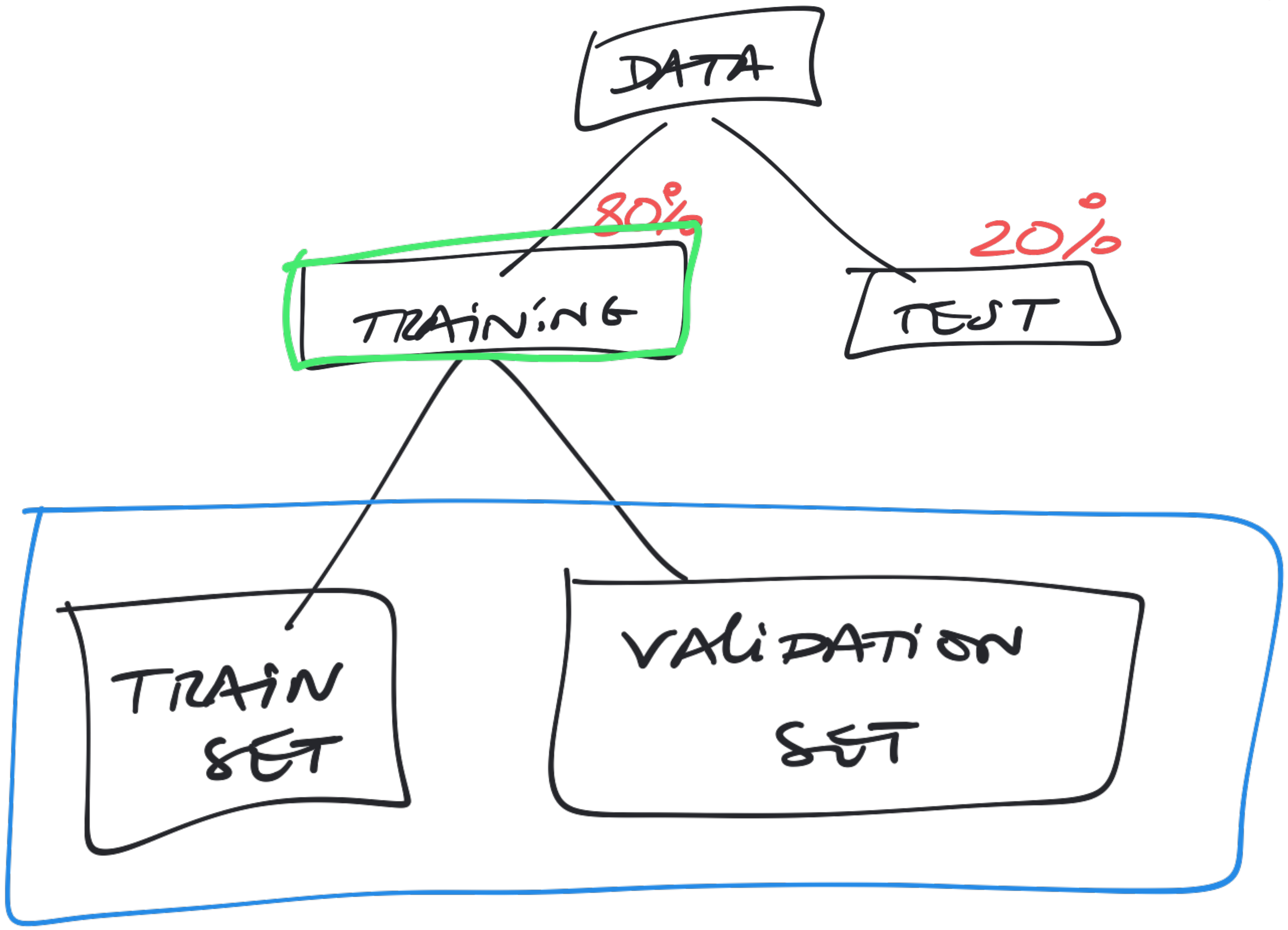
2. Cross - validation

3. Regularization

4. Low complexity — Apply Occam's Razor principle

Cross-Validation

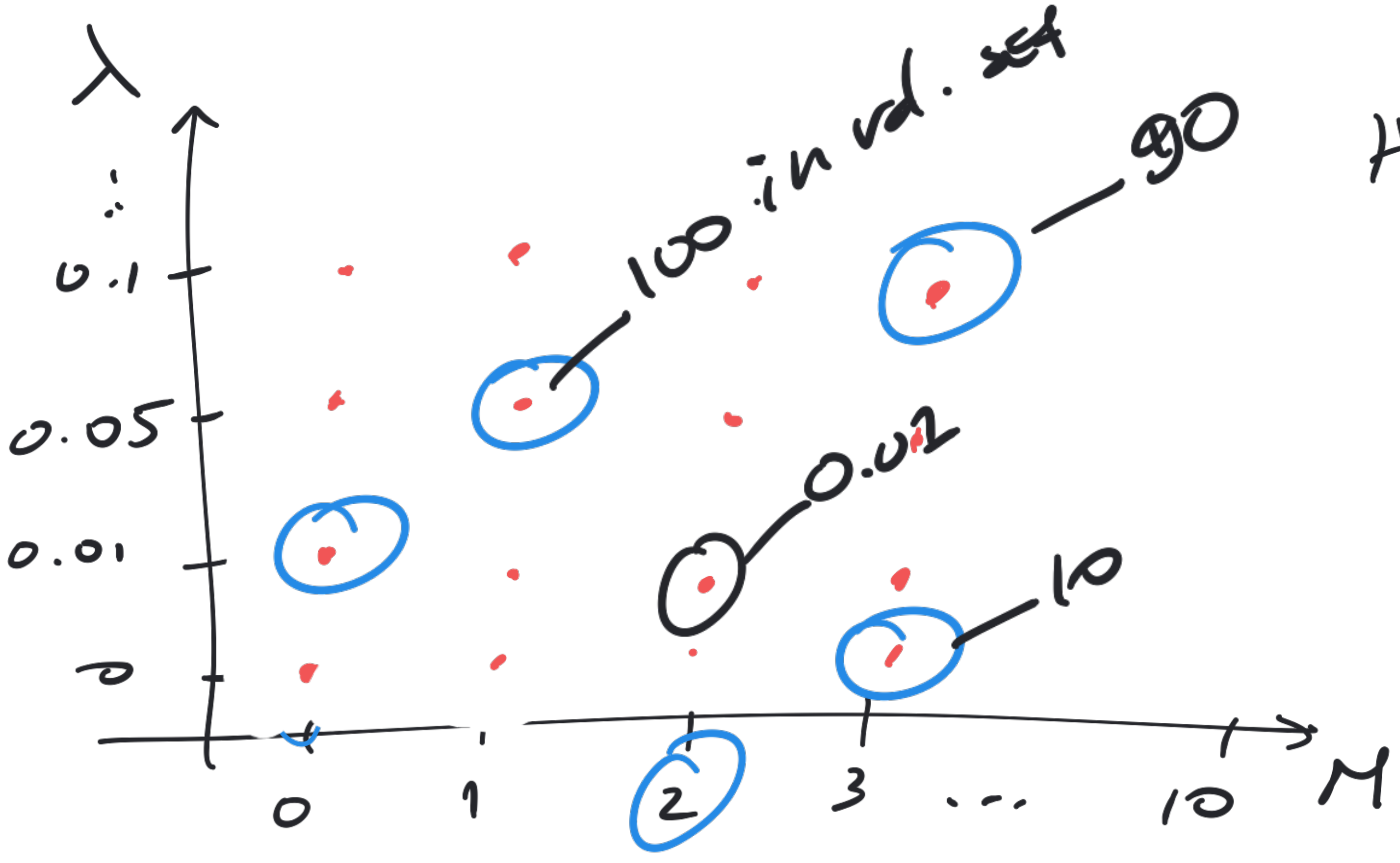
Helps us determine the best value for the hyperparameters using the training data.



Regularized Polynomial Regression

$$\{M, \lambda\} = \theta$$

Experimental Design via Cross-Validation



Hyperparameter space

Ⓐ Factorial Design. Experiment all possibilities.

Ⓑ One parameter at a time.
Requires more knowledge about problem/engineer regarding the model.

Ⓒ RANDOM SEARCH.

K-FOLD CV

1. DEFINE PERFORMANCE MEASURE. E.g. MSE, coefficient of determination (R^2), ...

2. PARTITION DATA

TRAIN SET (80%)

S_1	S_2	S_3	S_4

4-fold CV

(it usually a good idea to shuffle the data before partitioning it into folds)

3. ITERATE through several possible values for the hyperparameters (M and λ):

(*)

(*)

for model order M in some range: $M = \{2, 3, \dots, 10\}$

for regularizer term λ in some range: $\lambda = \{0, 0.01, 0.05, 0.1\}$

(1) TRAINING SET : $\{S_1, S_2, S_3\} \leftarrow$ MSE train = 50 $\quad w_1$

VAL. SET : $\{S_4\} \leftarrow$ MSE val. = 100

(2) TRAIN SET : $\{S_1, S_2, S_4\} \leftarrow$ MSE train = 0.5 $\quad w_2$

VAL. SET : $\{S_3\} \leftarrow$ MSE val = 1

(3) TRAIN SET : $\{S_1, S_3, S_4\} \leftarrow 2$

VAL. SET : $\{S_2\} \leftarrow 3$

(4) TRAIN SET : $\{S_2, S_3, S_4\} \leftarrow 10$

VAL. SET : $\{S_1\} \leftarrow 20$

(A) MSE AVERAGE in val. SET $= \dfrac{100+1+3+20}{4}$

train. SET $= \dfrac{50+0.5+2+10}{4}$

(B) MSE MAXIMUM in val. SET = 100

train SET = 50

(4) Pick set of hyperparameters that optimizes performance in validation set.

STRATIFIED CV

Partition the data such that prior probabilities for each class representation is preserved in all subsets.