Regularization

Ridge

$$R_w(w) = \sum_i w_i^2$$

Lasso

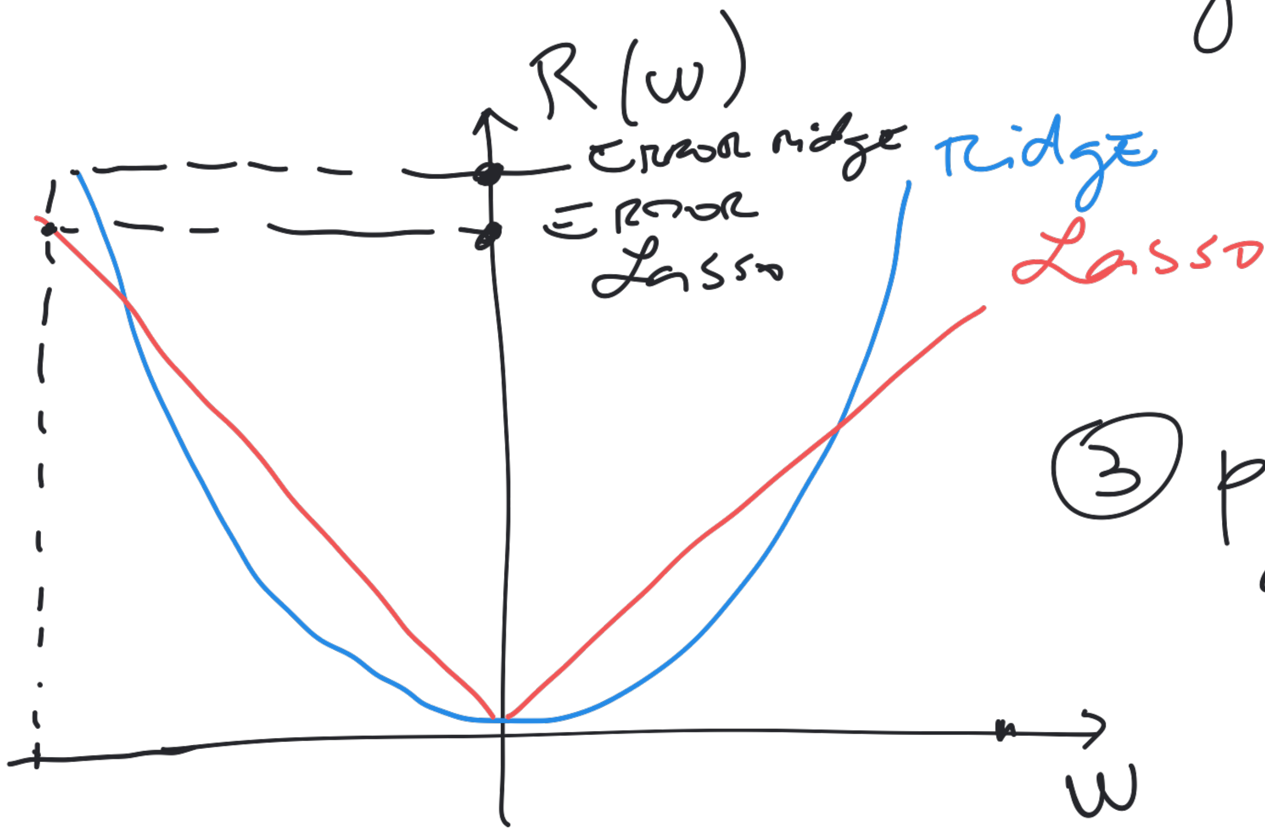
$$R_w(w) = \sum_i |w_i|$$

$$J(w) = \boxed{\frac{1}{2} \sum_n e_n^2} + \boxed{\frac{1}{2} \underline{\underline{\lambda}} \cdot R_w(w)}$$

Regularization weight

$$\boxed{\arg\min_w J(w)}$$

Ridge : ① Will penalize large weight values more Lasso



② Highly affected by outliers

③ prefers that coefficients are not zero but very small

Lasso : ① Less sensitive to outliers

② Makes weights ~~go~~ to zero much faster

③ prefers some coefficients exactly zero.

Using Lasso reg. we can perform <u>Feature Selection</u>

$$J(w) = J_E(x,w) + \lambda \cdot R_w(w)$$

$\lambda = 0$: only minimizing $J_E(x,w)$

$\lambda \to \infty$: disregards $J_E(x,w)$ and forces R_w to be small

FOR Regularized Polynomial Regression:

WE control:

1. Model
2. Cost function
3. Learning algorithm

Model order $\underline{\underline{M}}$ and reg. weight $\underline{\underline{\lambda}}$

$$J(w) = \frac{1}{2} \sum_{i=1}^{N} \underbrace{(t_i - y_i)}_{= e_i}^2 + \frac{1}{2} \cdot \lambda \cdot \sum_{j=0}^{M} w_j^2$$

$$= \frac{1}{2} \|t - Xw\|_2^2 + \frac{1}{2} \cdot \lambda \cdot \|w\|_2^2$$

If we have $N \gg M$: likely will also be an overdetermined system of eqs

\# samples

model order

$$\boxed{w^* = \left(X^T X + \lambda \cdot I\right)^{-1} \cdot X^T \cdot t}$$

regularized polynomial regression

→ identity matrix

$$\lambda \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & & & 1 \end{bmatrix} (M+1) \times (M+1)$$

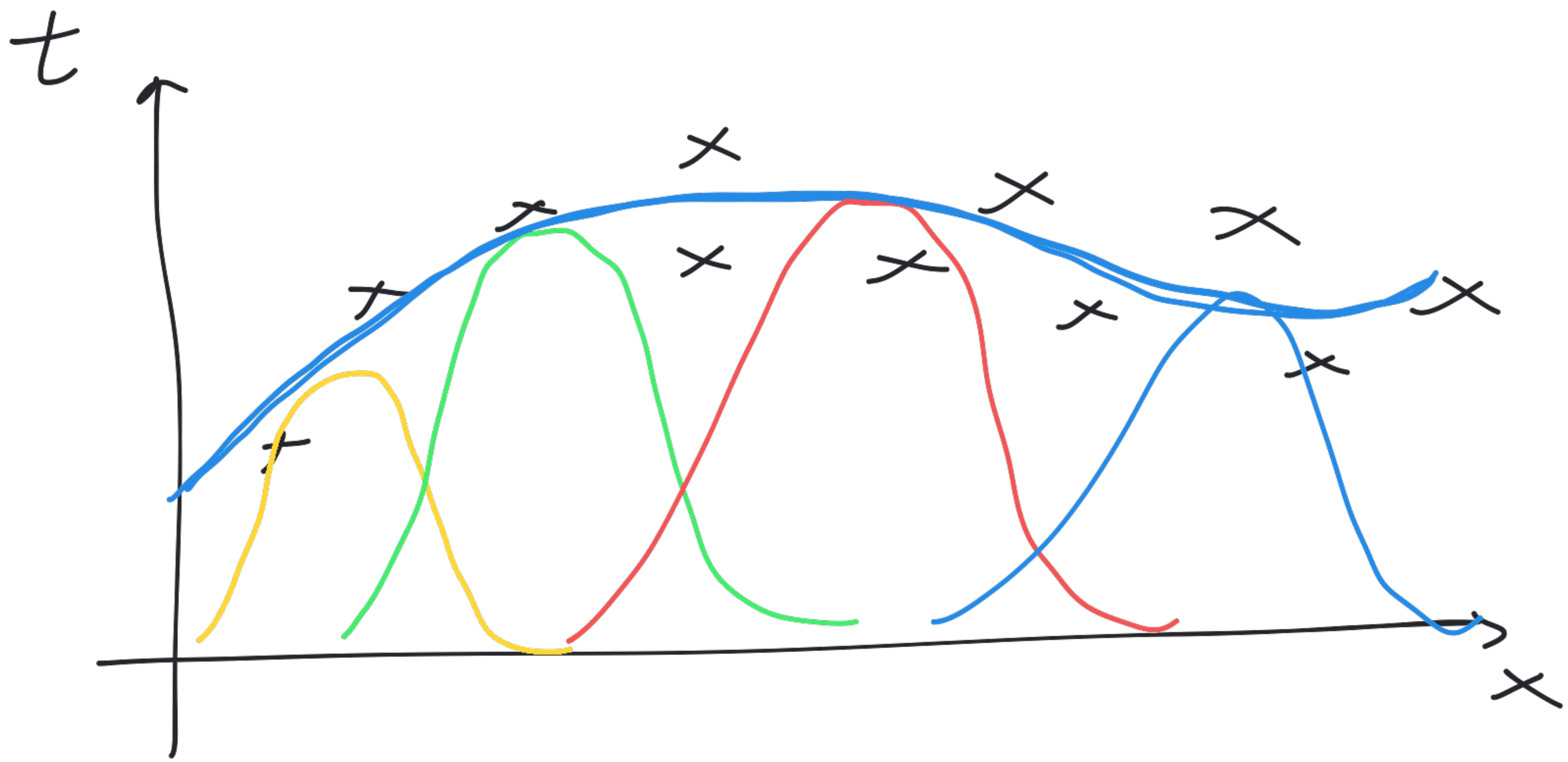
$$\boxed{w^* = \left(X^T X\right)^{-1} \cdot X^T \cdot t}$$

polynomial reg. w/out reg.

Regularization:

→ ~~diagonally loading~~ $X^T X$

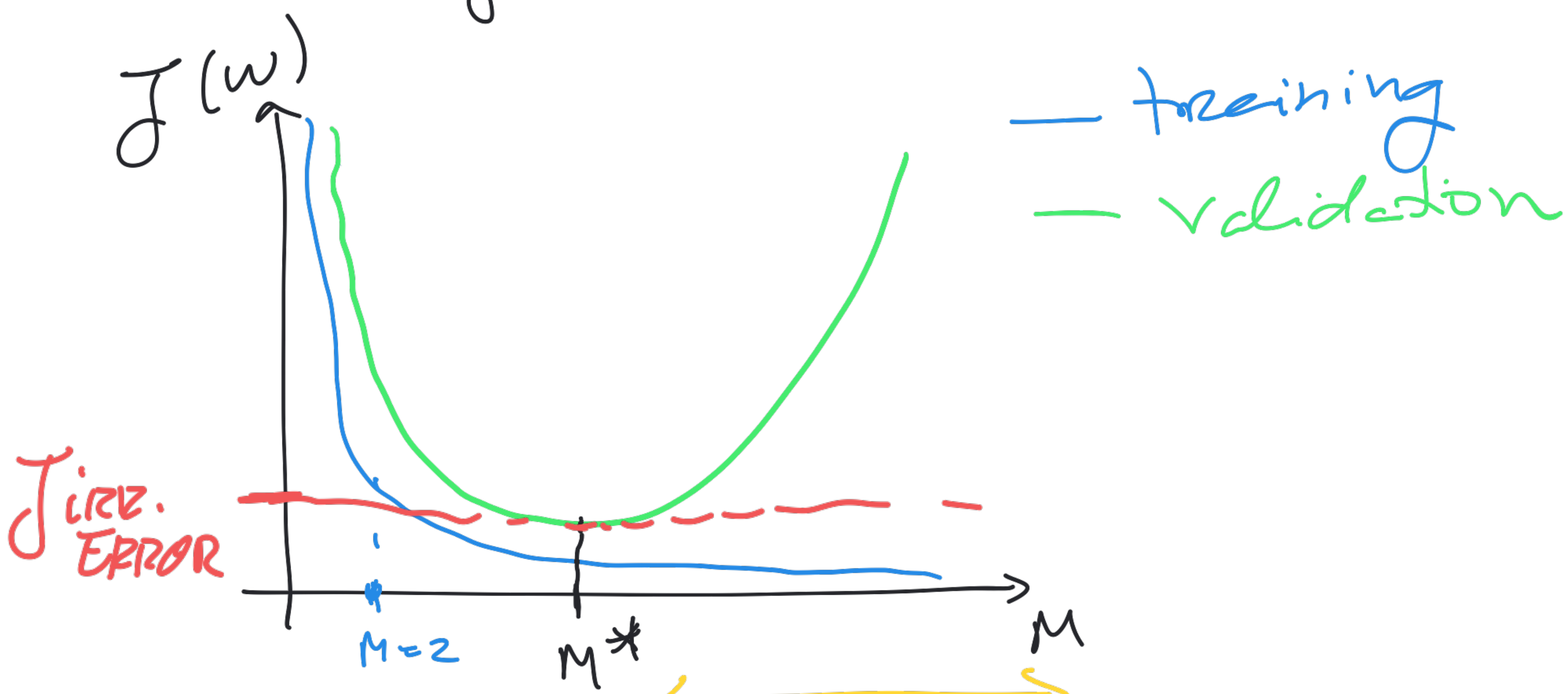
→ make such $X^T X$ is full rank



$$w_0 \cdot G + w_1 \cdot G + w_2 G + w_3 G$$

$$y = \sum_{j=0}^{M} w_j \cdot \phi_j(x)$$

- we are looking for the best combination of controllable parameters (M, λ):



underfitting

J_{train} will be high

J_{val} will be high

$\boxed{\text{high bias, low variance}}$

overfitting

J_{train} will be small

$J_{val} \gg J_{train}$

$\boxed{\text{low bias \& high variance}}$