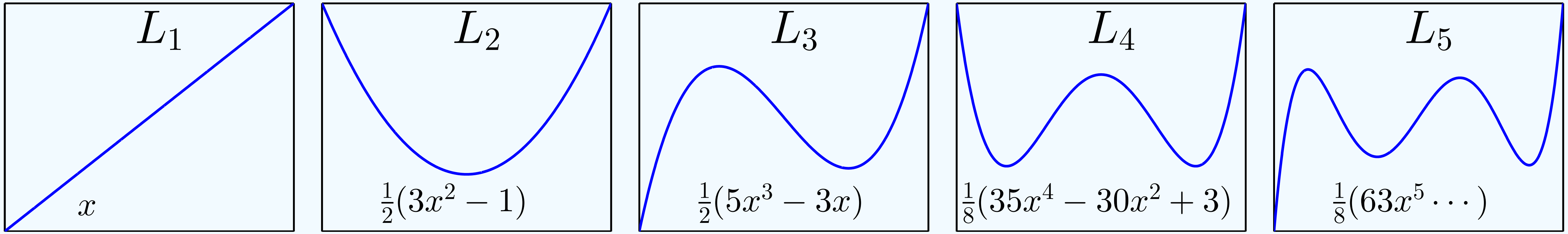
The polynomial model

$\mathcal{H}_Q$: polynomials of order Q

linear regression in $\mathcal{Z}$ space

$$\mathbf{z} = \begin{bmatrix} 1 \\ L_1(x) \\ \vdots \\ L_Q(x) \end{bmatrix} \qquad \mathcal{H}_Q = \left\{ \sum_{q=0}^{Q} w_q \, L_q(x) \right\}$$

Legendre polynomials:



Unconstrained solution

Given $(x_1, y_1), \cdots, (x_N, y_n) \longrightarrow (\mathbf{z}_1, y_1), \cdots, (\mathbf{z}_N, y_n)$

Minimize $E_{\text{in}}(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{w}^{\intercal}\mathbf{z}_n - y_n)^2$

Minimize $\frac{1}{N} (\mathrm{Z}\mathbf{w} - \mathbf{y})^{\intercal}(\mathrm{Z}\mathbf{w} - \mathbf{y})$

$$\mathbf{w}_{\text{lin}} = (\mathrm{Z}^{\intercal}\mathrm{Z})^{-1}\mathrm{Z}^{\intercal}\mathbf{y}$$

Constraining the weights

Hard constraint: $\mathcal{H}_2$ is constrained version of $\mathcal{H}_{10}$ with $w_q = 0$ for $q > 2$

Softer version: $$\sum_{q=0}^{Q} w_q^2 \le C$$ "**soft-order**" constraint

Minimize $\frac{1}{N}(\mathrm{Z}\mathbf{w} - \mathbf{y})^\intercal(\mathrm{Z}\mathbf{w} - \mathbf{y})$

subject to: $\mathbf{w}^\intercal\mathbf{w} \le C$

Solution: $\mathbf{w}_{\mathrm{reg}}$ instead of $\mathbf{w}_{\mathrm{lin}}$

Solving for $\mathbf{w}_{\text{reg}}$

Minimize $E_{\text{in}}(\mathbf{w}) = \frac{1}{N}(\mathrm{Z}\mathbf{w} - \mathbf{y})^{\mathsf{T}}(\mathrm{Z}\mathbf{w} - \mathbf{y})$

subject to: $\mathbf{w}^{\mathsf{T}}\mathbf{w} \leq C$

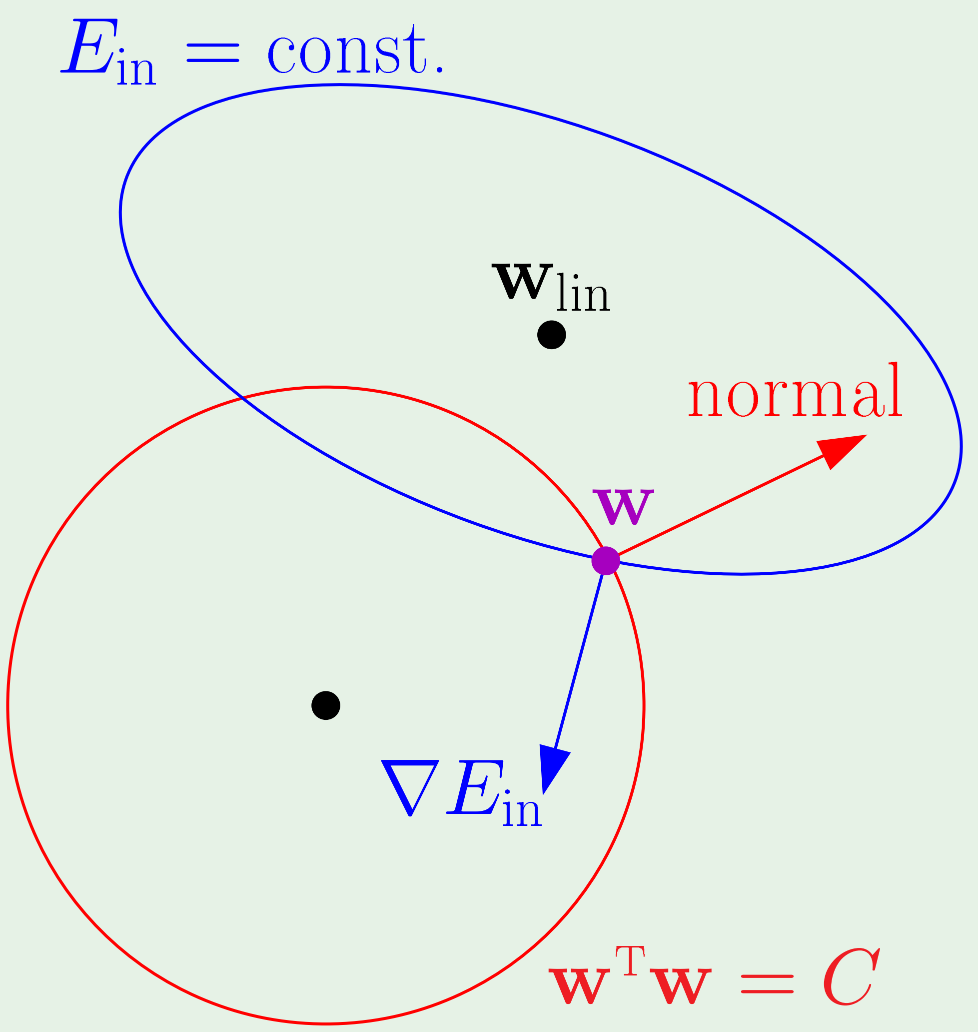
$$\nabla E_{\text{in}}(\mathbf{w}_{\text{reg}}) \propto -\mathbf{w}_{\text{reg}}$$

$$= -2\frac{\lambda}{N}\mathbf{w}_{\text{reg}}$$

$$\nabla E_{\text{in}}(\mathbf{w}_{\text{reg}}) + 2\frac{\lambda}{N}\mathbf{w}_{\text{reg}} = \mathbf{0}$$

Minimize $E_{\text{in}}(\mathbf{w}) + \frac{\lambda}{N}\mathbf{w}^{\mathsf{T}}\mathbf{w}$

$C \uparrow \quad \lambda \downarrow$



Augmented error

Minimizing $$E_{\text{aug}}(\mathbf{w}) = E_{\text{in}}(\mathbf{w}) + \frac{\lambda}{N}\mathbf{w}^{\intercal}\mathbf{w}$$

$$= \frac{1}{N}(\mathrm{Z}\mathbf{w} - \mathbf{y})^{\intercal}(\mathrm{Z}\mathbf{w} - \mathbf{y}) + \frac{\lambda}{N}\mathbf{w}^{\intercal}\mathbf{w}$$ unconditionally

— solves —

Minimizing $$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}(\mathrm{Z}\mathbf{w} - \mathbf{y})^{\intercal}(\mathrm{Z}\mathbf{w} - \mathbf{y})$$

subject to: $\mathbf{w}^{\intercal}\mathbf{w} \leq C$ $\longleftarrow$ VC formulation

The solution

Minimize $\quad E_{\text{aug}}(\mathbf{w}) \; = \; E_{\text{in}}(\mathbf{w}) + \frac{\lambda}{N}\mathbf{w}^{\mathsf{T}}\mathbf{w}$

$$= \; \frac{1}{N} \left((\mathrm{Z}\mathbf{w} - \mathbf{y})^{\mathsf{T}}(\mathrm{Z}\mathbf{w} - \mathbf{y}) \; + \; \lambda \; \mathbf{w}^{\mathsf{T}}\mathbf{w} \right)$$

$$\nabla E_{\text{aug}}(\mathbf{w}) \; = \; \mathbf{0} \qquad \Longrightarrow \qquad \mathrm{Z}^{\mathsf{T}}(\mathrm{Z}\mathbf{w} - \mathbf{y}) + \lambda\mathbf{w} = \mathbf{0}$$

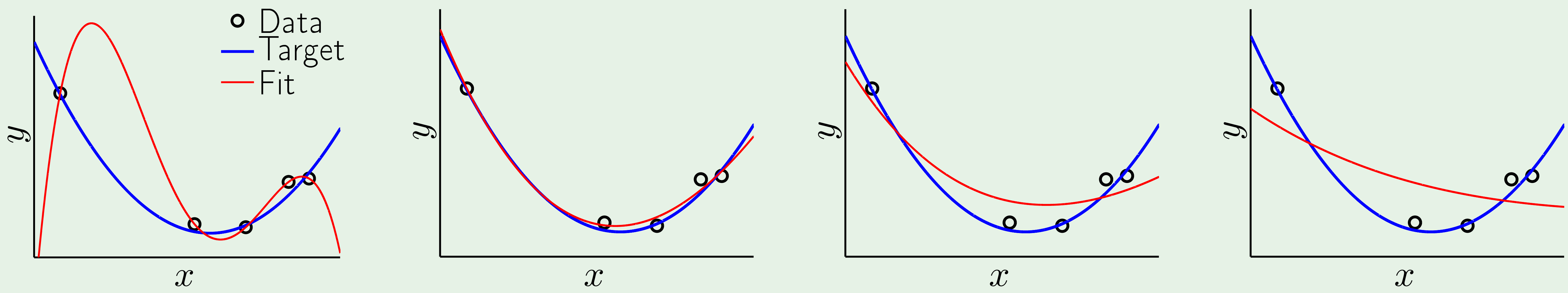
$$\boxed{\mathbf{w}_{\text{reg}} \; = (\mathrm{Z}^{\mathsf{T}}\mathrm{Z} + \lambda\mathrm{I})^{-1}\,\mathrm{Z}^{\mathsf{T}}\mathbf{y}}$$ (with regularization)

as opposed to $\quad \mathbf{w}_{\text{lin}} \; = (\mathrm{Z}^{\mathsf{T}}\mathrm{Z})^{-1}\mathrm{Z}^{\mathsf{T}}\mathbf{y}$ (without regularization)

The result

Minimizing $\quad E_{\text{in}}(\mathbf{w}) + \frac{\lambda}{N}\,\mathbf{w}^{\intercal}\mathbf{w} \quad$ for different λ's:

$\lambda = 0$ $\qquad$ $\lambda = 0.0001$ $\qquad$ $\lambda = 0.01$ $\qquad$ $\lambda = 1$



overfitting $\longrightarrow \quad \longrightarrow \quad \longrightarrow \quad \longrightarrow$ **underfitting**