General form of augmented error

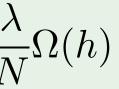
Calling the regularizer $\Omega = \Omega(h)$, we minimize

$$E_{\text{aug}}(h) = E_{\text{in}}(h) + \frac{\lambda}{N}$$

Rings a bell?

 $E_{\text{out}}(h) \leq E_{\text{in}}(h) + \Omega(\mathcal{H})$

 $E_{
m aug}$ is better than $E_{
m in}$ as a proxy for $E_{
m out}$



$\downarrow \downarrow$

16/21

Outline

- Regularization informal
- Regularization formal
- Weight decay
- Choosing a regularizer

17/21

The perfect regularizer Ω

Constraint in the 'direction' of the target function (going in circles \odot)

Guiding principle:

Direction of **smoother** or "simpler"

Chose a bad Ω ?

We still have λ !

18/21