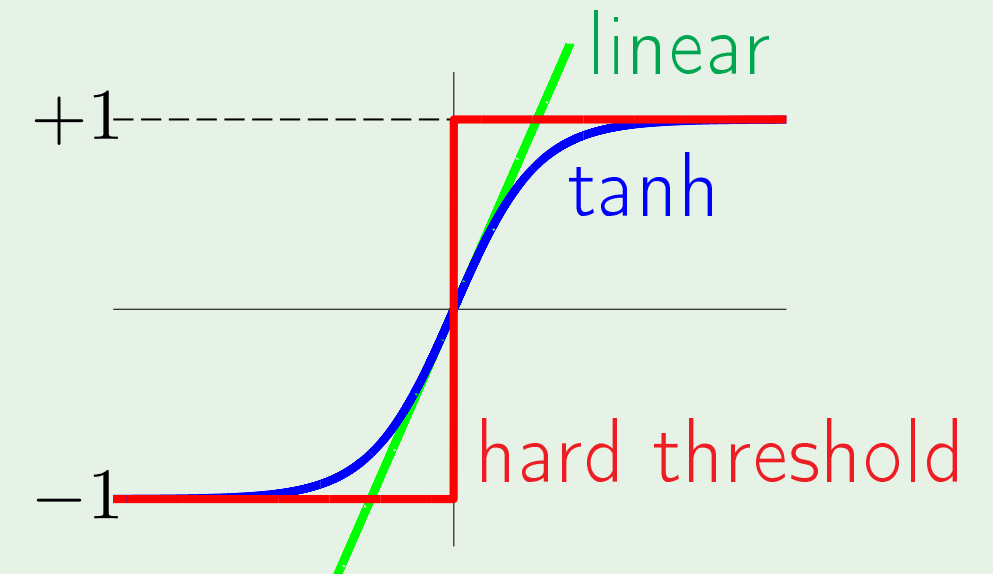


Neural-network regularizers

Weight decay: From linear to logical

Weight elimination:

Fewer weights \implies smaller VC dimension



Soft weight elimination:

$$\Omega(\mathbf{w}) = \sum_{i,j,l} \frac{\left(w_{ij}^{(l)}\right)^2}{\beta^2 + \left(w_{ij}^{(l)}\right)^2}$$

Early stopping as a regularizer

Regularization through the optimizer!

When to stop? **validation**

