

REGULARIZATION | SIMPLICITY

You can overcross! A model that is too complicated might fit to noise and then perform worse on the test data.

Regularization: don't trust examples too much. We want to make sure it can generalize.

↳ Avoids overfitting.

- Stop early (don't train too long)
- Penalize model complexity:

Empirical risk min: minimize $\text{Loss}(\text{Data}/\text{Model})$
Structural risk: minimize $\text{Loss}(\text{Data}/\text{Model}) + \text{Complexity}(\text{Model})$

How to define complexity (model)?

L2 regularization (Ridge):

- penalize sum (sq. weights)
- prefers flatter slopes, expects small weights normally distributed around zero

$$= \text{Loss}(\text{Data}/\text{Model}) + \lambda (w_1^2 + \dots + w_n^2)$$

↑
coefficient that scales complexity factor, needs tuning. = regularization rate
(lambda)

Avoids overfitting to training data!