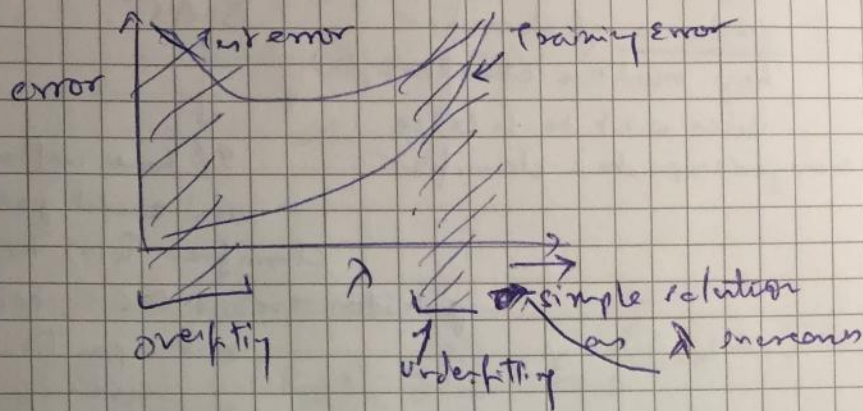


CS4780 L20 Regularization / Overfitting

→ Regularization

$$\min_w \sum l(x, y; w) + \lambda r(w) \quad \text{regularizer} = r(w)$$



→ Cross validation for λ search

→ Telescopic search $\lambda = 10^5, 10^4, 10^3, 10^2$

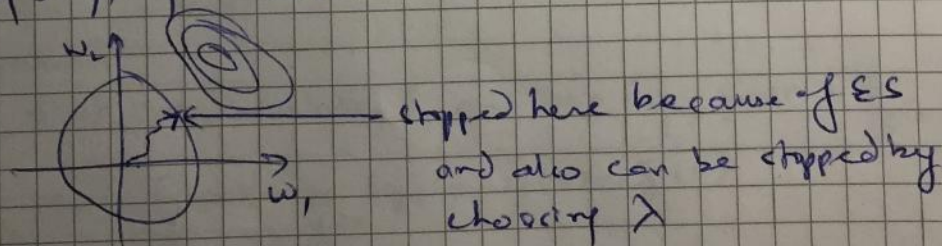
Search in this region.

May be then start overfitting on validation if validation is small.

→ k fold cross validation - k validation errors

$k = n$, n is size of data "Leave one out"

→ Early stopping - Connection between ES & Regularization



Other way As you take steps your ℓ_2 norm may grow, so cap the norm by ES.

