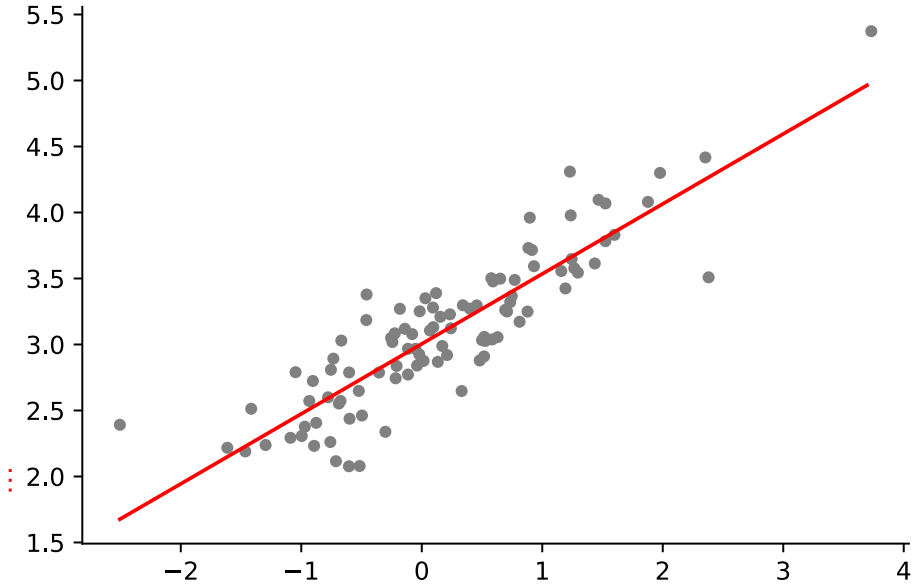
ODM Project 2021

Adversarial Training Set Selection for Regression

Linear Regression

Disease Progression
Housing Prices
Crop Yields
…



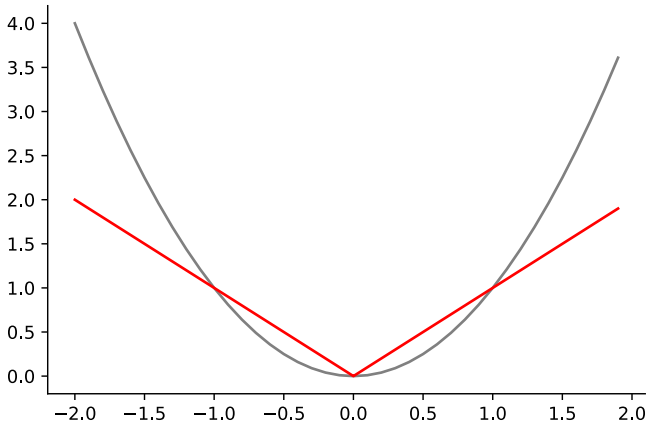
Blood Measurements
Number of Rooms
Rainfall/Sun
…

Loss Functions for Linear Regression

- $L(r) = r^2$
  - **M**ean **S**quared **E**rror
  - Gaussian:
    $p(x) \sim e^{-(x-\mu)^2/2\sigma^2}$
- $L(r) = |r|$:
  - **M**ean **A**bsolute **E**rror
  - Laplace
    $p(x) \sim e^{-|x-\mu|/b}$

$$y = \theta^T x + \xi$$

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} L(y_i - \theta^T x_i)$$



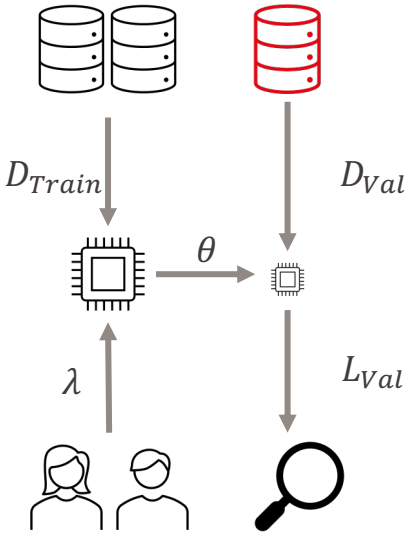
Regulaization

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y_i - \theta^T x_i) + \lambda R(\theta)$$

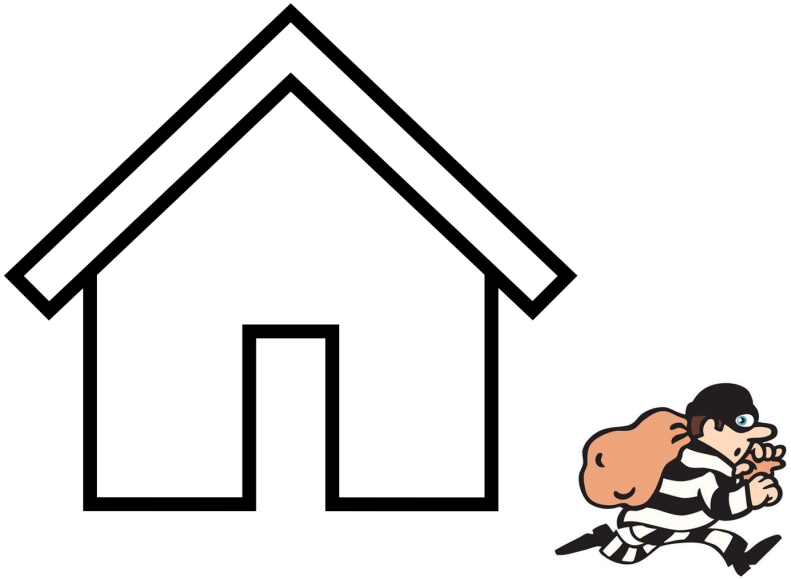
- Enforce Properties on θ:
  - Small Magnitude
  - Sparse (i.e., lots of zeros)

Train-Validation Split

- Problem: What should λ be?
- Solution:
  - Train on Part of your Data
    Training Set D_{Train}
  - Use the rest to find best λ
    Validation Set D_{Val}



Zero-Sum Games



$$\min_{x \in X} \max_{z \in Z} l(x, z)$$

What the best thief can steal: The most I can lose