Machine Learning

Regularization and Feature Selection

Fabio Vandin

November 22^{nd}, 2022

Feature Selection

In general, in machine learning one has to decide what to use as features (= input) for learning.

Even if somebody gives us a representation as a feature vector, maybe there is a "better" representation?

What is "better"?

Example

- features x_1, x_2, output y
- $x_1 \sim Uniform(-1, 1)$
- $y = x_1^2$
- $x_2 \sim y + Uniform(-0.01, 0.01)$

If we want to predict y, which feature is better: x_1 or x_2?

- x_1, because with x_1^2 you perfectly predict y. But what if you use x_1 as feature of a linear model? Prediction is not great!
- x_2, because with linear models we predict y exactly up to noise ($Uniform(-0.01, 0.01)$). But we could do better with x_1^2!

Example

- features x_1, x_2, output y
- $x_1 \sim \mathit{Uniform}(-1, 1)$
- $y = x_1^2$
- $x_2 \sim y + \mathit{Uniform}(-0.01, 0.01)$

If we want to predict y, which feature is better: x_1 or x_2?

No-free lunch...