27-04-2024

What is machine learning?

- Extract (learn) meaningful information from the examples (training) and answer the query for unseen examples
- Specific task
- How can we allow computer to extract meaningful information from the examples?
  - Design general rules
  - Development of algorithms
- Automate the process of inductive inference

Slide credit: Luxburg, 2020

Inductive bias

- Garcia effect, Garcia et. al. 1960
  - Rat has an inductive bias

Group-1

+

Sweet water

Group-2

+

drinking normal water

Sweet and normal water

drinking sweet water

Images: https://gemini.google.com/

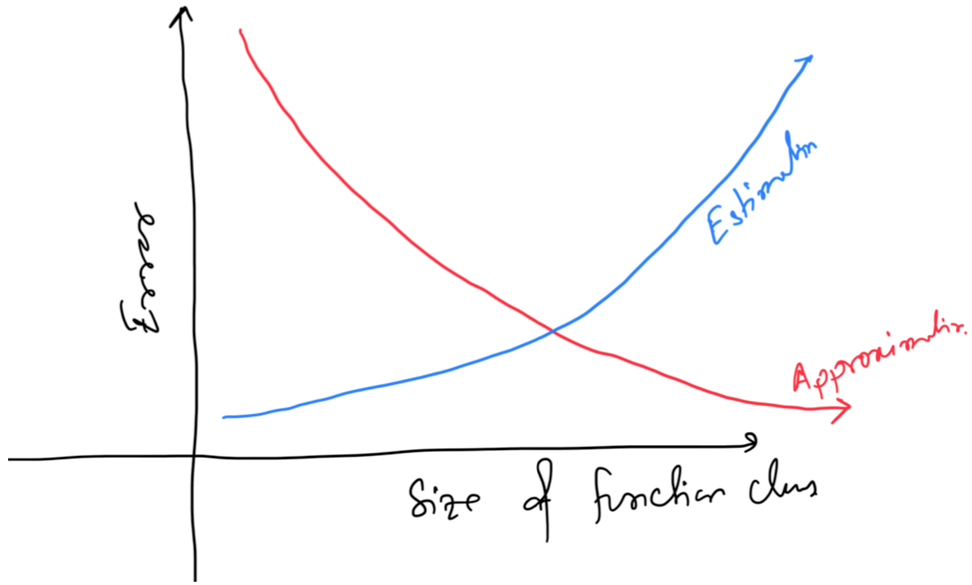
Empirical risk minimization (ERM)

- Given $X_1, X_2, \ldots, X_n$, $X_i \in \mathcal{X}$ are n data points and $Y_1, Y_2, \ldots, Y_n$, $Y_i \in \mathcal{Y}$ are their corresponding outputs
- If we know the data distribution, then we can use Bayes decision rule?
- But in most of the cases we don't know the data distribution!
  - We can compute ERM as: $R_n(f) := \frac{1}{n} \sum_{i=1}^{n} l(X_i, Y_i, f(X_i))$
  - Consider a set of functions $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$
    - $f_n := \arg\min_{f \in \mathcal{F}} R_n(f)$

Slide credit: Luxburg, 2020

Estimation Vs Approximation error

- Let's consider an example: least square regression problem
- Let $\hat{f}$ be the true best in $\mathcal{F}$:
  - $\hat{f} := \arg\min_{f \in \mathcal{F}} R(f)$
- Estimation error: $R(f_n) - R(\hat{f})$
  - Its a random variables. Why?
    - Depends on n
- Approximation error: $R(\hat{f}) - R(f^*)$
  - f^* is the Bayes/best over all functions (measurable)
- Size of the function class $\mathcal{F}$
  - Small ?
    - Undercutting
  - Large ?
    - Overfitting



Slide credit: Luxburg, 2020