CS60020: Foundations of Algorithm Design and Machine Learning

Sourangshu Bhattacharya
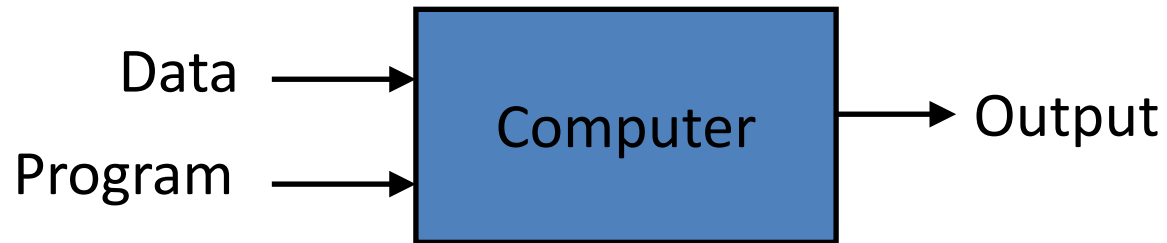
Algorithms

- An algorithm is an unambiguous specification of how to solve a class of problems.

- Example: Euclid's algorithm for finding the greatest common divisor.
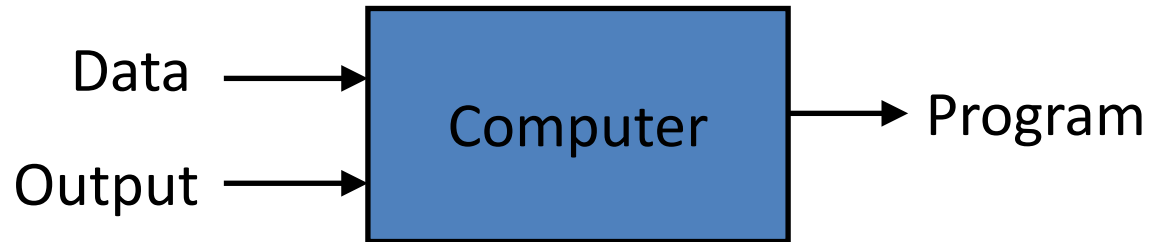
- Important Aspects:
  - Analysis
  - Design

Machine Learning

- Machine learning is a field of computer science that gives computers the ability to learn [from data] without being explicitly programmed.

- Example: Bayesian classifier for automatically filtering email spams.

- Aspects:
  - Modeling
  - Inference and learning

Traditional Programming

Data $\rightarrow$ **Computer** $\rightarrow$ Output

Program $\rightarrow$ **Computer**

Machine Learning

Data $\rightarrow$ **Computer** $\rightarrow$ Program

Output $\rightarrow$ **Computer**

Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging
- [Your favorite area]

ORGANISATION

Venue

- Classroom: CSE – 119
- Slots:
    - Monday (10:00 – 11:00 am)
    - Wednesday (8:00 – 10:00 am)
    - Thursday (10:00 – 11:00 am)
- Website:
  http://cse.iitkgp.ac.in/~sourangshu/coursefiles/cs60020_18S.html
- Moodle:
  https://10.5.18.110/moodle/

Pre-requisites

- Basic Programming and Data Structures.
- Mathematics: functions, matrix algebra, optimization.
- Discrete Maths: Graphs, Trees, etc.

- Logical thinking and Hard work !!

Teaching Assistants

- Rijula Kar
- Chandan Misra
- Saptarshi Misra
- Manaar Alam

Evaluation

- Grades:
  - TA (Assignments and class tests): 40
  - Mid-sem: 30
  - End-sem: 30

- Assignments: 6 – 10 (small ones).
- Class test: 2.

SYLLABUS

Algorithm Design

- Introduction to the design and analysis of efficient algorithms.

- Time and space complexity and order notation.

- Trees. Lists. Hashing.

- Graph and Basic graph algorithms: Definitions and representation, Reachability, Shortest path.

- Divide and Conquer and Dynamic programming

- NP-completeness- basic notions.

- Randomized algorithms and approximation algorithms.

- Sub-linear algorithms, streaming algorithms, sampling

Machine Learning

- Concept learning. Hypothesis space. Inductive Bias. Learnability. Underfitting and overfitting.

- Feature Selection, Dimension Reduction

- SVM and introduction to kernel methods.

- Unsupervised and semi-supervised learning. Expectation maximization. Mixture of Gaussians.

- Active learning, Learning with Imbalanced Data. Anomaly detection.

- Ensemble methods.

- Introduction to graphical models

- Introduction to Deep Learning