Fundamental Algorithm Techniques

Problem Set #10

Yessengaliyeva Gulnazym

December 2025

Problem 1: P, NP, NP-complete, NP-hard

We classify each group of problems into its corresponding complexity class. Each line below contains problems belonging to the same class.

1. **Class: P**

   Finding the maximum, linear search, shortest path in an unweighted graph, and matrix multiplication are all solvable in polynomial time.

2. **Class: P**

   Sorting a list, Dijkstra's algorithm with non-negative weights, BFS, DFS, merge sort, and quicksort all admit polynomial-time algorithms.

3. **Class: NP-complete**

   Sudoku (in its generalised $n \times n$ form) belongs to NP, since a solution can be verified in polynomial time, and it is NP-hard via known reductions. Thus, it is NP-complete.

4. **Class: NP-complete**

   Graph 3-coloring is a classical NP-complete problem. Scheduling with conflicts can be reduced to graph coloring, and therefore belongs to the same class.

5. **Class: NP-complete**

   The Traveling Salesperson Problem (decision version), Hamiltonian Cycle, and Clique are all canonical NP-complete problems.

6. **Class: NP**

   Factoring large integers lies in NP since a factorisation can be verified efficiently. It is widely used in cryptography. It is not known whether this problem belongs to P.

7. **Class: Undecidable**

   The Halting Problem and the Busy Beaver problem are undecidable, meaning that no algorithm can solve these problems for all possible inputs.

Problem 2: Introduction to Bayes' Theorem

A deadly disease affects 0.1% of the population:

$$P(D) = 0.001.$$

The medical test has:

- sensitivity $P(+ \mid D) = 0.99$,
- specificity $P(- \mid \neg D) = 0.99$.

Therefore,

$$P(+ \mid \neg D) = 0.01, \qquad P(\neg D) = 0.999.$$

Using Bayes' theorem, the probability that a patient actually has the disease given a positive test result is:

$$P(D \mid +) = \frac{P(+ \mid D)\, P(D)}{P(+ \mid D)\, P(D) + P(+ \mid \neg D)\, P(\neg D)}.$$

Substituting the values:

$$P(D \mid +) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.01 \cdot 0.999} = \frac{0.00099}{0.01098} \approx 0.09.$$

Answer. The probability that the patient actually has the disease is approximately **9%**.

Explanation. Although the test is highly accurate, the disease itself is very rare. As a result, most positive test results come from false positives among healthy individuals. This phenomenon is known as the *base rate fallacy*.

Problem 3: Introduction to Shannon Entropy

We consider three coins with different probabilities of landing heads. The Shannon entropy is defined as:

$$H(X) = -\sum_i p_i \log_2 p_i.$$

Coin A: Fair coin ($P(H) = 0.5$)

$$H = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) = 1 \text{ bit}.$$

A fair coin has maximum uncertainty, so each toss carries exactly one bit of information.

Coin B: Biased coin ($P(H) = 0.99$)

$$H = -(0.99 \log_2 0.99 + 0.01 \log_2 0.01) \approx 0.08 \text{ bits}.$$

Since the outcome is highly predictable, the amount of information gained from each toss is very small.

Coin C: Biased coin ($P(H) = 0.01$)

This coin has the same entropy as Coin B:

$$H \approx 0.08 \text{ bits.}$$

Although a rare event is more surprising when it happens, such events occur so infrequently that the average information content remains low.

Conclusion. A fair coin is worth 1 bit of information because it has maximal uncertainty, while a strongly biased coin is worth much less, as its outcome can be predicted with high confidence.