Problem Set #10

Fundamental Algorithm Techniques

Amirtayeva Ainur

This report contains solutions to all tasks from **Problem Set #10**

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

We classify each line exactly as required.

Line 1

find max, linear search, shortest path in unweighted graph, matrix multiplication **Class: P**. All problems admit polynomial-time solutions.

Line 2

sorting, Dijkstra (non-negative weights), BFS, DFS, merge sort, quicksort **Class: P**. These are classical efficient algorithms.

Line 3

sudoku **Class: NP-complete**. General sudoku solving is a canonical NP-complete problem.

Line 4

3-coloring, scheduling with conflicts **Class: NP-complete**. Both are known NP-complete decision problems.

Line 5

Traveling Salesperson (decision version), Hamiltonian Cycle, Clique **Class: NP-complete**. These belong to the core set of NP-complete problems.

Line 6

cryptography, factoring large integers **Class: NP-hard / unknown**. Factoring is not known to be in P nor NP-complete; cryptographic hardness assumptions rely on difficulty of such problems.

Line 7

Halting Problem, busy beaver **Class: Undecidable**. No algorithm can solve these problems in general.

Summary:

- Lines 1–2: P
- Lines 3–5: NP-complete
- Line 6: NP-hard / unknown
- Line 7: Undecidable

Problem 2. Bayes Rule and the 9% Result

A disease affects 0.1% of people:

$$P(D) = 0.001.$$

The test is 99% accurate:

$$P(+|D) = 0.99, \quad P(-|\neg D) = 0.99,$$

so the false positive rate is

$$P(+|\neg D) = 0.01.$$

We apply Bayes' Theorem:

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

Substitute values:

$$P(D|+) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.01 \cdot 0.999} = \frac{0.00099}{0.01098} \approx 0.09 = 9\%.$$

Explanation

Although the test is very accurate, the disease is extremely rare. Most positive tests come from the much larger group of healthy people, so the chance of truly having the disease is only 9%.

Problem 3. Shannon Entropy of Three Coins

Entropy:

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

Each coin has two outcomes: Heads and Tails.

Coin A: $P(H) = 0.5$

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

A fair coin is maximally unpredictable.

Coin B: $P(H) = 0.99$

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

The outcome is almost always Heads, so it carries very little information.

Coin C: $P(H) = 0.01$

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

Again, the coin is almost deterministic.

Conclusion

A fair coin is worth exactly 1 bit because the uncertainty is maximal. A highly biased coin provides little new information per flip, therefore its entropy is close to 0.08 bits.