Problem Set #10 – Solutions

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

Line 1 (find max, linear search, shortest path in unweighted graph, matrix multiplication): Class P. All have polynomial-time algorithms.

Line 2 (sorting, Dijkstra with non-negative weights, BFS, DFS, merge sort, quicksort): Class P. All run in polynomial time.

Line 3 (Sudoku): NP-complete (in NP and NP-hard in its general form).

Line 4 (3-coloring, scheduling with conflicts): NP-complete.

Line 5 (TSP, Hamiltonian Cycle, Clique): NP-complete (decision versions).

Line 6 (cryptography, factoring large integers): Factoring is in NP, but not known to be in P and not known to be NP-complete.

Line 7 (Halting Problem, Busy Beaver): Undecidable.

Problem 2. Bayes' Theorem (Medical Test)

Given: $P(D) = 0.001$ (prevalence), $P(+|D) = 0.99$ (sensitivity), $P(-|\text{not } D) = 0.99$ (specificity), so $P(+|\text{not } D) = 0.01$.

Bayes' formula:

$$P(D|+) = (P(+|D) \cdot P(D)) / (P(+|D) \cdot P(D) + P(+|\text{not } D) \cdot P(\text{not } D))$$

Substitute values:

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

It is low because the disease is rare, so false positives among healthy people dominate the positive results.

Problem 3. Shannon Entropy

Entropy for a discrete random variable with outcomes i and probabilities p_i is:

$$H(X) = - \Sigma\, p_i \log_2(p_i)$$

For a coin with P(Heads)=p and P(Tails)=1-p:

$$H(p) = - p \log_2(p) - (1-p) \log_2(1-p)$$

Coin A ($p = 0.5$): $H = -0.5 \cdot \log_2(0.5) - 0.5 \cdot \log_2(0.5) = 1$ bit.

Coin B ($p = 0.99$): $H \approx -0.99 \cdot \log2(0.99) - 0.01 \cdot \log2(0.01) \approx 0.08$ bits.

Coin C ($p = 0.01$): same entropy as Coin B (symmetry), $H \approx 0.08$ bits.

A fair coin is maximally uncertain (1 bit). A 99% biased coin is highly predictable, so the average information per flip is much smaller.