TD 1
Proofs, computability and distances

Version du September 20, 2020

Exercice 1 – Inductive proofs

1. (**Easy induction to warm you up**) Prove by induction that splitting the n squares of a chocolate bar requires to break up the bar $n-1$ times (whatever the order you choose for breaking the grooves).
2. (**Be careful with the baseline cases**) Show that any natural number greater than or equal to 8 can be written as $3a+5b$, with $a, b \in \mathbb{N}$ (e.g., $19 = 3+3+3+5+5$).
3. (**Structural induction**) Let a tree be defined as follows:
   — An isolated *node* is a tree, and is its own root. The degree of this node is 0.
   — A new tree can be constructed from a list $A_1, A_2, \ldots, A_n$ as follows: we create a new node N that will be the root of the new tree, and we link it with edges to the root of each A_i. The degree of N is n.

   a) Show that each tree has one node more than its number of edges.

   b) We consider a tree, all of whose nodes have degree 0 or 2. Show that such a tree contains k nodes of degree 2 if it contains $k+1$ nodes of degree 0.
4. What is the link between question 1 and question 3?

Exercice 2 – Computability

1. Are the following sets (a) recursively enumerable, (b) recursive?
   — The set of all prime numbers.
   — The set of polynomials with natural number coefficients.
   — The empty set.
   — The set of all programs which do not loop forever.
   — The set of all programs which end in less than 10 seconds.
2. (**A non recursively enumerable language**) Let Σ be some alphabet, and let $m_0, m_1, m_2, \ldots$ be the sequence of words of $\Sigma^\star$ sorted in alphabetical order. If $\Sigma = \{a, b\}$ for instance, the alphabetical order is $\epsilon, a, b, aa, ab, ba, bb, \ldots$

   Also, let $A_0, A_1, A_2, \ldots$ be the set of algorithms that recognize the recursively enumerable languages of $\Sigma^\star$ similarly sorted (their code (in your favorite programming language) is sorted in alphabetical order).

   Prove (by contradiction) that the language L defined as

   $$L = \{m_i \mid A_i \text{ does not recognize } m_i\}$$

   is not recursively enumerable.

Exercice 3 – Prefix distance

Let u, v and w be three words using the alphabet Σ. We remind the reader that $\mathrm{lcp}(u, v)$ stands for the longest common prefix of u and v.

1. Justify that $|\mathrm{lcp}\big(\mathrm{lcp}(u,v), \mathrm{lcp}(v,w)\big)| \leq |\mathrm{lcp}(u,w)|$.
2. Justify that $|\mathrm{lcp}(u,v)| + |\mathrm{lcp}(v,w)| \leq \min(|\mathrm{lcp}(u,v)|, |\mathrm{lcp}(v,w)|) + |v|$
3. Justify that $|\mathrm{lcp}\big(\mathrm{lcp}(u,v), \mathrm{lcp}(v,w)\big)| = \min(|\mathrm{lcp}(u,v)|, |\mathrm{lcp}(v,w)|)$.
4. Deduce that $|\mathrm{lcp}(u,v)| + |\mathrm{lcp}(v,w)| \leq |\mathrm{lcp}(u,w)| + |v|$.
5. Show that $d_p(u,v) = |uv| - 2\,|\mathrm{lcp}(u,v)|$ is a distance.
   (previous questions will be useful to prove the triangular inequality).

Exercice 4 – Edit distance

The edit distance (also called Levenshtein distance) between u and v is the smallest number of elementary edit operations (insertion or deletion of a symbol) allowing to transform u into v.

1. Propose a recursive definition of this distance.

$$d_L(u,v) = ?$$

2. Prove that it is indeed a distance.