**Short question examples (Mid-terms and part of finals)**

Q1.1 What is the maximum degree (the number of edges from a vertex) in the prefix tree for the string “ACCATACCACACCA”? What is the 2nd largest degree?

Q1.2 For the Hamiltonian Problem with *n* vertices, suppose I give you a “witness” (certificate), how would you check that it is correct? And, what is the computational complexity of your procedure?

Q1.4 Draw the transition table for a Finite State Automaton that will parse all occurrences of the string “AA[C|A]”, where [C|A] means C or A.

Q1.6. For a nondeterministic Finite State Automata, why is the maximum possible number of machines that could be running at the same time dependent on the number of states *k* and not the length of the patter? Would it ever be useful to have more than *k* machines running simultaneously?

Q1.8. Some string P has BWT string ‘ACGTG$A’. Please give the original string P.

Q1.9. Compute the distance between vectors x = (3, 4, 1) and y = (2, 1, 6) using the radial basis kernel function: .

**Finals question example**

I have used next generation sequencing of the rRNA locus from DNA/RNA extracted from a swab taken from a patient’s skin. That is, we extracted all of the DNA/RNA from a pool of microbial cells and sequenced the rRNA genes en mass in a procedure called “metagenomics”. Metagenomic sequencing does not attempt to isolate individual microbial cells, but simply “cracks open” every cell and sequences the pool of DNA/RNA from this extraction.

rRNA is a conserved gene that is easy to sequence and can be used to identify microbial species and also estimate their evolutionary relationships. I didn’t get a whole lot of sequences but I was able to get 1,000 100bp reads of presumed microbial rRNA sequences. Assume that there is a database of 10,000 different microbial species and their rRNA sequences.

Q6.1 Describe a method to identify the different microbial species/strains present in the skin sample. Also describe how we might quantify the relative abundance of each species/strains.

Q6.2 Given the data generated in Q6.1, I wish to know whether microbiome can be used to predict patients with irritable bowel syndrome (IBS). I obtained microbiome measurements from 100 IBS patients and 100 normal controls. Suggest a method to predict IBS from microbiomes. Make sure you discuss feature space, learning machine, training data, test data, and identifying key species/strains that are associated with IBS.