Part I. An incomplete list of important concepts

Pairwise Alignment
* edit distance, gap penalty, local/global alignments
* log likelihood ratio, Bayesian rules, frequentist inference, null hypothesis, P-value, E-value.
* blosum matrices, blosum 62.
* linear space alignment

Multiple Alignment
* relative entropy score, SP score, approximation algorithm,

Homology search
* spaced seed, multiple spaced seed

HMM
* HMM, transition probability, emission probability
* HMM parameter estimation, higher order HMM
* codon bias, start/stop codons
* prokaryote gene prediction

Phylogeny
* Perfect phylogeny
* Parsimony, distance based method, UPGMA, neighbor joining, quartet.
* Use compression (Kolmogorov complexity) for distance between whole genomes.

Suffix Tree and Array
* suffix tree can be constructed in linear time
* use suffix tree and array to do pattern matching

Mass Spectrometry
* MS, MS/MS, ionizer, mass analyzer, MALDI, ESI, TOF, Quadrupole, Orbitrap
* protease, enzyme, trypsin, amino acid residue mass, peptide mass, ion,
* precursor ion, fragment ion, b and y ions.
* de novo peptide sequencing, database search method for peptide identification from MS/MS,
* false discovery rate, target-decoy, decoy fusion
* PTM, fixed PTM, variable PTM
Part II. Algorithms to know well
* For all problems/algorithms studied in the class, you should have the basic knowledge about the
time/space complexity and the best available algorithm.
* The following is a short list of those you should know well: understand, memorize, prove
time/space complexity and correctness, and slightly modify them to solve similar problems.
* For all the dynamic programming (DP) algorithms, know the recurrence relation, as well as
how to initialize the table, and how to do backtracking.

Alignment
* pairwise alignment DP (linear gap, arbitrary gap, affine gap penalty)
* local alignment
* linear space global alignment (know that it can be extended for local alignment)

Homology search
* use spaced seed to find hits
* compute the sensitivity of a spaced seed

Multiple Alignment
* dynamic programming for multiple alignment
* merging two multiple alignments together: \((s1,s2,s3)+(s3,s4,s5)\Rightarrow(s1,s2,s3,s4,s5)\)
* heuristic algorithm
* ratio-2 approximation algorithm for SP score

HMM
* find the optimal path of states

Phylogeny
* perfect phylogeny for binary case
* ancestor reconstruction to minimize total number of mutations.
* UPGMA

Suffix Tree and Suffix Array
* quadratic time algorithm for constructing Suffix Tree,
* finding the longest common substring of two strings in linear time
* linear time algorithm for constructing suffix array

Mass Spectrometry
* de novo sequencing algorithm with only y-ions (or only b-ions)
* target-decoy approach for result quality control and its difference to decoy fusion.
* linear time finding the nearby peaks in two sorted peak list