Part I. A partial list of important concepts/knowledge

Pairwise Alignment

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

Multiple Alignment
* SP score, approximation algorithm

Homology search
* spaced seed, multiple spaced seed
* hit and extend, 2 hits

Suffix Tree, Suffix Array, FM-Index
* pattern matching with each data structure
* suffix tree: longest common substring, maximum unique matching
* suffix array: prefix doubling, skew algorithm
* FM-index: BWT, reconstruction original text, backward query. No need to remember the space saving techniques.

Genome Sequencing
* Sanger sequencing, NGS

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

HMM
* HMM, transition probability, emission probability
* HMM parameter estimation, higher order HMM
* codon bias, start/stop codons
* prokaryote gene prediction
* Eukaryote gene prediction - combine NN with HHM

Protein Structure Prediction
* Primary, secondary, tertiary structures
* Gibbs free energy
* Torsion angles
* Contact map

* Rosetta (overall procedure, no need to memorize all details)
* AlphaFold (overall structure, no need to memorize all details)

Deep Learning in Bioinformatics

* CNN, RNN, Transformer (self and cross attentions)

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. This list may be useful for your study but may not be complete.
* 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)=>(s1,s2,s3,s4,s5)
* heuristic algorithm
* ratio-2 approximation algorithm for SP score

HMM

* find the optimal path of states

Suffix Tree, Suffix Array, FM-Index

* quadratic time algorithm for constructing Suffix Tree,
* finding the longest common substring of two strings in linear time
* prefix doubling, skew algorithm for suffix sorting
* reconstruction original text from BWT, backward query in FM-index.

Mass Spectrometry

* de novo sequencing algorithm
* target-decoy approach for result quality control