INSTRUCTIONS:

1. Submit your solution to Gradescope by the due time; no late submissions will be accepted.
2. Type your solution (except figures; figures can be hand-drawn and then scanned); no hand-written solutions will be accepted.

Problem 1 (10 points).

Let T be the dynamic programming table when computing the edit distance between two strings. Prove that along any diagonal of T the values are non-decreasing.

Problem 2 (10 points).

Let S be a set of reads, each of which is of length at most L. Prove that the shortest common superstring (SCS) of S does not contain a repeat of length $2L-1$.

Problem 3 (10 points).

Given two strings s and t, design an algorithm to find the optimal suffix-prefix alignment (Lecture 19). Specifically, the algorithm should find a suffix of s, a prefix of t, and an alignment between the suffix and the prefix such that its score is maximized. You are given the score matrix $score(a,b)$, $a,b \in \Sigma$, which gives the score when letter a is aligned to letter b. You may assume the unit gap score, i.e., the score of a single gap "-" is a fixed given parameter g, $g < 0$. Your algorithm should run in $O(|s| \cdot |t|)$ time.

Problem 4 (10 points).

You run an ice cream business, and you want to place some advertisements in your local newspaper. There are two kinds of ads you can run, Type-C and Type-W, and you've noticed that Type-C works best on cold days (by promoting the good taste of your ice cream) and Type-W works best on warm days (by mentioning how cold and refreshing your ice cream is). Depending on the weather and which ad you run, you see a certain amount of increased profit that day: on a cold day the profit will be \$75 if running a Type-C ad and \$50 if running a Type-W ad, on that day; on a warm day the profit will be \$50 if running a Type-C ad and \$100 if running a Type-W ad, on that day. You have committed to running an ad every day. The cost of placing either a Type-C or Type-W ad is \$10 per day. But the newspaper charges you a fee of \$25 every time you change which ad you are running. You are given a (perfectly correct) weather prediction for the next n days. Design a dynamic programming algorithm to select which ad to run on each of the next n days to maximize your total profit. For examples, for an input being WWWCCCWCWCWCW, where C/W indicates a cold/warm day, you should output WWWCCCWWWWWWW, where C/W indicates running a Type-C/typw-W ad, with a total profit of \$895. Your algorithm should run in $O(n)$ time.

Problem 5 (10 points).

Consider the (global) edit distance problem. Often there are multiple optimal alignments (i.e., all have the same, minimized edit distance). Given two sequences s and t, design an algorithm to compute the number of distinct optimal alignments. Your algorithm should run in $O(|s| \cdot |t|)$ time. *Hint:* the number of distinct optimal alignments can be obtained by computing the number of optimal traceback paths.

Problem 6 (10 points).

You are given an undirected graph $G = (V,E)$ and an integer k; assume that $|E| \geq k$. You aim to remove k edges from G such that in the resulting graph the number of connected components is maximized. Formulate this problem as an ILP. *Hint:* the main idea can be borrowed from Lecture 21.