

# Problem Set #5

제출기한: 2022/4/15 11:59 AM

\* 모든 문제의 답을 “pset05\_영문이름\_학번.m”으로 이름붙인 하나의 파일에 스크립트로 작성하여 제출할 것. 문  
제별로 적절한 주석을 사용하여 구분할 것. 한 문제의 답은 스크립트, 함수, 또는 둘을 혼합하여 작성할 수 있다.  
과제 안에서 재사용 가능한 함수를 적극적으로 만드는 것도 좋다.

\* 좋은 코드의 조건을 두루 고려하라. 좋은 코드는 목적을 정확히 달성해야 하고, 효율적으로 작동해야 하며, 함수들은 일반성과 재사용성을 갖추고 있어야 하며, 변수와 함수의 이름은 내용을 잘 반영해야 한다.

\* 조건문과 반복문은 필요한 경우에 적절히 사용하라. 반드시 필요한 경우가 아니라면 쓰지 않는 편이 좋다.

0. 네번째 과제의 3번 문제를 조건문과 반복문을 사용하지 않고 해결해보시오. 배열의 기초를 활용하는 것으로 충분하다. 2주차 강의노트 12페이지를 참고하라. A([3 2; 1 5])?

pset04.mat 파일의 sample\_in\_grid와 sample\_to\_species를 사용하십시오. sample\_in\_grid는 어떤 지역에서 조사한 구획별 거주 생명체를 나타낸다. 이 지역을  $m \times n$ 의 바둑판 구획으로 나누어 행렬로 나타내고, 이 행렬의 원소는 각 구획에서 발견한 생명체를 특징에 따라 분류한 숫자이다. 원소가 0인 구획은 생명체가 없음을 나타낸다. 이후, 임시로 분류했던 생명체 종류  $L$ 가지가 실은  $N$ 가지의 종이라는 사실을 알게 되었다. 분류된 종류와 최종 종의 대응 관계가  $L \times 2$  행렬 형태의 표 sample\_to\_species에 정리되어 있다. 이들을 활용하여,  $m \times n$  지역에서 각 구획별로 살고 있는 생명체 종을 표시하십시오. 즉,  $m \times n$  행렬 sample\_in\_grid의 0이 아닌 원소를 종을 나타내는 새로운 숫자로 변환하면 된다.

이 문제는 배점이 없고 제출하지 않아도 괜찮지만, 스스로 실력을 향상시키고 싶다면 도전하기 바라고 제출하면 다른 문제와 같이 피드백을 해줄 예정입니다.

1. 임의의  $k$ 에 대해 파스칼의 삼각형을  $k$ 행까지 화면에 출력하라. 파스칼의 삼각형의  $k$ 행( $k \geq 3$ )은  $k$ 개의 숫자를 가지며 첫 숫자와 마지막 숫자가 1이고 그 사이의 숫자들은 위 행의 인접한 두 수의 합으로 주어진다.  $k=6$ 일 때 아래 그림과 같이 출력되어야 한다. 1행의 1이 최대한 가운데 오고 나머지 행들도 보기 좋게 정렬되도록 해보자. (60점)

$$\begin{array}{ccccccc}
 & & & & 1 & & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
 1 & & 5 & & 10 & & 10 & & 5 & & 1
 \end{array}$$
[illegible]

3. 아래 bioinformatics 문제에 대해 각각 함수를 작성하고 테스트해보라.

(1) Counting Point Mutation (<https://rosalind.info/problems/hamm/>)

같은 길이의 문자열 두 개 s와 t가 주어질 때, 둘 사이의 “Hamming 거리”는 같은 자리에 있는 문자가 다른 경우의 개수로 정의된다. 즉, Hamming 거리는 두 개의 유전자 서열에 대해 한점 돌연변이(point mutation)의 개수와 같다. 같은 길이의 문자열 입력 s와 t 사이의 Hamming 거리를 측정하여 반환하는 함수를 작성하라. (15점)

Input

s='GAGCCTACTAACGGGAT'

t='CATCGTAATGACGGCCT'

Output

7

(2) Transitions and Transversions (<https://rosalind.info/problems/tran/>)

한점 돌연변이 중에서, purine 끼리 (A-G) pyrimidine 끼리 (C-T) 의 돌연변이를 transition(전이), purine과 pyrimidine 사이의 돌연변이를 transversion(변위)라고 부른다. (문제 링크에서 Certain Point Mutations are More Common 부분을 클릭하여 전이와 변위의 관계도 참고.)

두 개의 string이 주어졌을 때, transition / transversion의 비율을 구하는 함수를 작성하라. (25점)

Input

s='GCAACGCACAACGAAAACCCTTAGGGACTGGATTATTTTCGTGATCGTTGTAGTTATTGGAAGTACGGGC  
ATCAACCCAGTT'

t='TTATCTGACAAAGAAAGCCGTCAACGGCTGGATAATTTTCGCGATCGTGCTGGTTACTGGCGGTACGAGT  
GTTCCCTTTGGGT'

Output

1.2143