1 Exercise 1

Find all optimal secondary structures of the RNA sequence AAUACUCCGUUGCAGCAU with the following crude energy minimization algorithm. Starting from the slides' algorithm, use the following initialisation:

$$j + 5 > i => E(i, j) = 100, i > j,$$

and bond energy: $-4, 0, 4$, for Watson-Crick bonds, GU, and all other possible pairs respectively. Implement your algorithm in Matlab, R, Python or other convenient system; submit your code. Print the filled-in table E. Draw (by hand) all optimal folds, show the bonds, and each corresponding backtrack path.

2 Solution

In this exercise we want to fold the RNA sequence AAUACUCCGUUGCAGCAU. Having as given the costs of the different combinations of letters, we just need to initialize the values of the matrix with 100,

```
size= len(seq)
matrix= np.ones(shape=(size,size))*100
```

The function `score()` will fill the table with the crude energy minimization algorithm with the given energies for every bond.

Using the Crude algorithm :

$$W(row, col) = min \begin{cases} W(row, col-1) \\ W(row+1, col) \\ W(row+1, col-1) + damageofpair(seq_{row}, seq_{col}) \\ min_k(W(k, col) + W(row, k-1) : row + 1 < k < col) \end{cases} \tag{1}$$

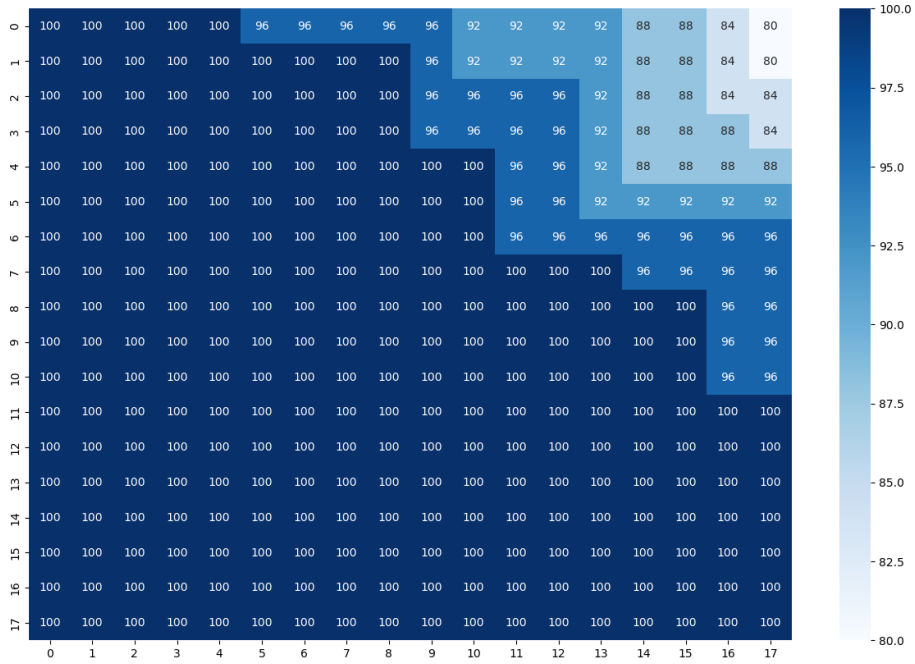
We can calculate the minimum energy for every bond. At this point, it is important to note, that the Crude algorithm will use the function `score()` in order to fill the matrix from the bottom to the top, starting from the last row and first column position.

Finally it find the folding that minimizes the cost with the functions.

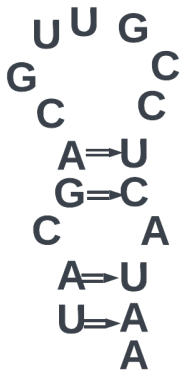
```
for row in range(size-1,-1,-1):
  for col in range(row+5, size):
    let1= seq[col]
    let2= seq[row]
    damage= score(let1,let2)
    temp1= matrix[row,col-1]
    temp2= matrix[row+1,col]
```

```
temp3= matrix[row+1,col-1]+ damage
tempk=[]
for k in range(row+2,col):
   tempk.append(matrix[k,col]+matrix[row,k-1])
matrix[row,col]=min(temp1,temp2,temp3,min(tempk))
```

Finally I ended up with the matrix.



And with the following optimal folding.


U U G
G C
C C
A U
G C
C A
A U
U A
A