**Natural Language Processing – Exercise 3**

:

Given Data:

Hidden states – {H, L}





Given sequence: **S = ACCGTGCA**

Starting state = H.

Viterbi Algorithm:

K=0:

We know that the starting state is H, and there are no back pointers yet.

K=1: A nucleotide was emitted

K=2: C nucleotide was emitted

K=3: C nucleotide was emitted

K=4: G nucleotide was emitted

K=5: T nucleotide was emitted

K=6: G nucleotide was emitted

K=7: C nucleotide was emitted

K=8: A nucleotide was emitted

As we can see the best end state is L as shown in K=8:

We will follow the back pointers and reach to the best state-sequence:

**Best sequence probability calculation:**

:

Pseudo-code for the Four-gram Viterbi Algorithm:

**Input:**

* An integer .
* Parameters and .

**Definitions:**

* : Set of possible tags.
* .
* .
* : Set of possible words.

**Initialization:**

1. Define (Base probability).
2. For .

**Algorithm:**

1. Iterate over positions to :
   * For each tag in :
     + For each tag in :
       - For each tag in :
         * Compute the maximum probability:
         * Store the corresponding in a back-pointer .
2. Final Step (n+1):
   * Set to:
   * Track back-pointers for the best sequence.
3. Return:
   1. Use the back-pointers to reconstruct the optimal tag sequence .
   2. Return the sequence and the maximum probability.