Assignment 3

1. Biological (50%, 35% for grads)

Construct, **if possible**, perfect phylogenetic trees for the two sets of the following aligned DNA sequences. Choose the binary attributes to be gaps (insertions **or** deletions).

```
A: RPCVCP___VLRQAAQ__QVLQRQIIQGPQQLRRLF_AA
B: RPCACP___VLRQVVQ__QALQRQIIQGPQQLRRLF_AA
C: KPCLCPKQAAVKQAAH__QQLYQGQLQGPKQVRRAFRLL
D: KPCVCPRQLVLRQAAHLAQQLYQGQ____RQVRRAF_VA
E: KPCVCPRQLVLRQAAH__QQLYQGQ____RQVRRLF_AA
```

```
A: RPCVCP___VLRQAAQLAQVLQRQIIQGPQQ___LF_AA
B: RPCACP___VLRQVVQKAQALQRQI____QQLRRLF_AA
C: KPCLCP___AVKQAAHLAQQLYQGQLQGPKQ___AFRLL
D: KPCVCPRQLVLRQAAH__QQLYQGQIQGPRQVRRAF_VA
E: KPCVCP___VLRQAAHLAQQLYQGQ_____RQVRRLF_AA
```

2. Non-biological (50%, 35% for grads)

John lives in Toronto. He is always in one of 2 moods: he is either sad or happy. When it rains, he is sad 3 times out of 4; when it is sunny, he is sad 1 time out of 6. We know that the sunny day follows the rainy day in Toronto with probability 2/5; the sunny day is followed by the rainy day with probability 5/6.

We talk with John every day on Skype using a webcam for 7 days. On the first day, he told us that it rains in Toronto and he looked sad.

What was the probable weather in Toronto on day 7, if we recorded the mood observation sequence as follows:

Sad, Sad, Happy, Happy, Happy, Sad, Happy?

3. For grad students (30%, and the bonus as usual for undergrads)

HMM Parameters			
	L	F	e(6)
B	0.50	0.50	
L	0.67	0.33	0.50
F	0.17	0.83	0.17

To avoid suspicion, the dealer in the occasionally dishonest casino (with the fair and loaded dies; see table above) keeps every die for at least 5 tosses, regardless whether it is fair or loaded. Describe the corresponding HMM. Explain how to modify the Viterbi algorithm for this case. Compare the decoding of the following sequence by two models: first – regular HMM with given parameters, second – constrained HMM of a smart dealer. You can use log scores (since actual probabilities when multiplied become too small) and you can modify the Excel tool presented in class.

Sequence: 1266656466166636