EX2 Report – Shon Cohen

1. Viterbi algorithm with first sequence from yeastGenes.fasta and p = 0.003. We can see 13 appearances of the motif.

```
CGGGAGAACTGGTTATAGGATTTTTTTTGTTTTTTTGTGTTTTAGAATTA
BBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBBBBBBBBBMMM
```

```
GATTGAATTATAAAAAGAAGAACAAAAGGGTATCGTATTGAAAATAAATT
MMMMBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
```

```
GTCTCGCCAAACTGGTAACAATGTTTTCAGCTAGAACAATAAGAAAAGAA
BBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBB
```

```
GAGAAGGTAAAAAAAAAGGTGATAACTCCGTAGGAATTGAGGAATTGAGT
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
```

```
ATGCAGAATCAGAATAAAGGCTGACTTTCAAAAAAAGGTTGTATTACAAT
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBB
```

```
TGCAGGTTTTCGATAAAAGAGACCCTATTCTCATCTACTACTGCTAAACT
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMM
```

```
TCGAGATATTTTCGAATTTTTCAGTCTTTTCTTTTTTTTTTTTCGCATTA
MBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
```

```
GTTCAGAACCCTAAAGAATGGTAAACATTCTATGGATAACCCGGAGAGTG
BBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBBBMM
```

```
AGTTTCTTAAAGACCTAGTTTTATTTTAAGGGTTTTAACTCAATCTTGAT
MMMMMBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBMMMMMMMBBB
```

```
GTTTTCATTGTGTACCCTAAAGAAAGTTTAAGAATAGCCCTAACTGTTAC
BBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBMMMM
```

CTTTTGAAATAAAATAAGGGGAAGGTCAAAAAGCTATTGTTCTATTGTTA
MMMBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBB

TGAAACATTGTCTCAAAGAGTAAGAATAACACAAATTGATGGCAGTTTTT
BBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBBBBBB

TACGTAGTCCAGTAGTTGTCCAGGTACAA
BBBBBBBBBBBBBBBBBBBBBBBBBBBBB

2. **Forward algorithm** – the log likelihood of the first sequence in yeastGenes.fasta: -869.610408405

3. **Posterior decoding** vs Viterbi:

CGGGAGAACTGGTTATAGGATTTTTTTTGTTTTTTTGTGTTTTAGAATTA
BBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBBBBBBBBBMMM
00000000000000000011110000000000000000000000000000

GATTGAATTATAAAAAGAAGAACAAAAGGGTATCGTATTGAAAATAAATT
MMMMBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
00000000000000000000000000000000000000000000000000

GTCTCGCCAAACTGGTAACAATGTTTTCAGCTAGAACAATAAGAAAAGAA
BBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBB
00000000000000000000000000000000000000000000000000

GAGAAGGTAAAAAAAAAGGTGATAACTCCGTAGGAATTGAGGAATTGAGT
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
00000000000000000000000000000000000000000000000000

```
ATGCAGAATCAGAATAAAGGCTGACTTTCAAAAAAAGGTTGTATTACAAT
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBB
00000000000000000000000000000000001111111000000000

TGCAGGTTTTCGATAAAAGAGACCCTATTCTCATCTACTACTGCTAAACT
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMM
00000000000000000000000000000000000000000000111111

TCGAGATATTTTCGAATTTTTCAGTCTTTTCTTTTTTTTTTTTCGCATTA
MBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
10000000000000000000000000000000000000000000000000

GTTCAGAACCCTAAAGAATGGTAAACATTCTATGGATAACCCGGAGAGTG
BBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBBBBMM
00000000000000000000100111000000000000000000000011

AGTTTCTTAAAGACCTAGTTTTATTTTAAGGGTTTTAACTCAATCTTGAT
MMMMMBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBMMMMMMMBBB
11111000000000000000000000000000000000000000000000

GTTTTCATTGTGTACCCTAAAGAAAGTTTAAGAATAGCCCTAACTGTTAC
BBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBMMMM
00000000000000000000000000000000000000000000000000

CTTTTGAAATAAAATAAGGGGAAGGTCAAAAAGCTATTGTTCTATTGTTA
MMMBBBBBBBBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBB
00000000000000000000000000000000011000000000000000

TGAAACATTGTCTCAAAGAGTAAGAATAACACAAATTGATGGCAGTTTTT
BBBBBBBBBBBBBBBBBBBBBMMMMMMMBBBBBBBBBBBBBBBBBBBBBBBB
00000000000000000001000000000000000000000000000000
```

TACGTAGTCCAGTAGTTGTCCAGGTACAA

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

000000000000000000000000000000

4. Motif learning: as we can see, the model converged successfully.