**Department of Computing and Information Systems**

**COMP 90016**

Workshop 5

This workshop will investigate the use of hidden Markov Models (HMMs) to locate CpG islands in DNA sequences following slides 15-27 from the lectures.

1. Use the following sequence to derive the state transition matrix for a HMM-based CpG island detector (CpG islands in yellow).

GGTTCCGCTCCCACCGCGCCGCGTTCGGCCGCGTT

1. Use the above estimates and the brute-force algorithm from slide 17 to annotate the sequence CGCGGCAAAGC

.

1. Check that your result is correct using the Viterbi (dynamic programming) algorithm.
2. Read the file test.data. Use your algorithm to classify each of the sequences. What is the accuracy of your model? How could it be improved?