

Computational Genomics

Week 9 Workshop

Viterbi Algorithm with CNV detection

Recap

- HMM
 - Start probability
 - Emission probability
 - Transition probability
- Viterbi algorithm
 - Probability of being in state L after observing x_{i+1} is:
 - $v_L(i+1) = e_L(x_{i+1}) * \max(v_K(i) * a_{KL})$
 - $O(nm^2)$

Recap

- Viterbi algorithm and HMM
 - Start probability
 - Emission probability
 - Transition probability
- Discuss with your neighbours, in the context of CpG island detection,
 - what does each probability mean?
 - what should each probability table look like?

What if the context becomes CNV detection?

- CNV detection
- ..by read depth

CNV detection

- Redefine the probabilities for CNV detection. Discuss with your neighbours.
- Hint: reference HMM on CpG island detection.

CNV detection

- States: copy number
- Emissions: read depth
- Start probabilities: probabilities of *first state*
- Emission probabilities: probabilities of *state -- emission*
- Transition probabilities: probabilities of *state -- state*

Implementation of Viterbi on CNV detection

- Assume flat start probability (equal for all states)
- Emission probability table in file: *emissions.txt*
- Transition probability table in file: *transitions.txt*
- Read depth of bins in file: *bins.txt*