Assignment-5

Guided by:- PROF. SAURABH GUPTA

TOPIC- Hidden Markov Model

SUMMARAY

Hidden Markov models (HMMs), named after the Russian mathematician Andrey Andreyevich Markov, who developed much of relevant statistical theory, are introduced and studied in the early 1970s. They were first used in speech recognition and have been successfully applied to the analysis of biological sequences since late 1980s. Nowadays, they are considered as a specific form of dynamic Bayesian networks, which are based on the theory of Bayes. HMMs are statistical models to capture hidden information from observable sequential symbols (e.g., a nucleotidic sequence). They have many applications in sequence analysis, in particular to predict exons and introns in genomic DNA, identify functional motifs (domains) in proteins (profile HMM), align two sequences (pair HMM). In a HMM, the system being modelled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters. A good HMM accurately models the real world source of the observed real data and has the ability to simulate the source. A lot of Machine Learning techniques are based on HMMs have been successfully applied to problems including speech recognition, optical character recognition, computational biology and they have become a fundamental tool in bioinformatics: for their robust statistical foundation, conceptual simplicity and malleability, they are adapted fit diverse classification problems. In Computational Biology, a hidden Markov model (HMM) is a statistical approach that is frequently used for modelling biological sequences. In applying it, a sequence is modelled as an output of a discrete stochastic process, which progresses through a series of states that are 'hidden' from the observer. Each such hidden state emits a symbol representing an elementary unit of the modelled data, for example, in case of a protein sequence, an amino acid. In the following sections, we first introduce the concepts of Hidden Markov Model as

a particular type of probabilistic model in a Bayesian framework; then, we describe some important aspects of modelling Hidden Markov Models in order to solve real problems, giving particular emphasis in its use in biological context. To show the potentiality of these statistical approaches, we present the stochastic modelling of an HMM, defining first the model architecture and then the learning and operating algorithms. In this work we illustrate, as example, applications in computational biology and bioinformatics and, in particular, the attention is on the problem to find regions of DNA that are methylated or un-methylated (CpG-islands finding)