CE956: Statistical Learning
Department of Computer Engineering
Sharif University of Technology
Spring 2019: Room CE204, Sat. & Mon.: 13:30-15:00

Quiz 03 (35 Points) – (March-16-2019)

Solution

Non-Parametric Models: (5 points each)

1. Briefly explain what is the Dirichlet Distribution and Dirichlet Process.
   - (Please refer to the class notes and reference papers). The Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a vector α of positive reals. It is a multivariate generalization of the beta distribution. The Dirichlet distribution is the conjugate prior of the categorical distribution and multinomial distribution. A Dirichlet process is a probability distribution whose range is itself a set of probability distributions.

2. Why Dirichlet Process is important in the context of Bayesian Nonparametric Models.
   - (Please refer to the class notes and reference papers). A Dirichlet process is often used in Bayesian inference to describe the prior knowledge about the distribution of random variables, referring to how likely it is that the random variables are distributed according to one or another particular distribution.

3. What is the difference between CRP and IBP?
   - (Please refer to the class notes and reference papers). The Chinese restaurant process is a discrete-time stochastic process, analogous to seating customers at tables in a Chinese restaurant with an infinite number of circular tables, each with infinite capacity. The Indian buffet process (IBP) is a stochastic process defining a probability distribution over sparse binary matrices with a finite number of rows and an infinite number of columns.

4. What is the relation between distributions over the binary matrices and CRP and IBP?
   - (Please refer to the class notes and reference papers). We may use binary matrix representation for clustering. In this representation, rows are data points, and columns are clusters. Since each data point is assigned to one and only one cluster, rows sum to one. In this context, the Chinese restaurant process (CRP) is the distribution on partitions of the data induced by a Dirichlet Process Mixture (DPM) where the number of columns is countably infinite. Thus, we can think of the CRP as a distribution on such binary matrices (hard membership). We may think of a more general distribution on binary matrices (soft membership) where rows are data points, and columns are latent features. We can think of infinite binary matrices where each data point can now have multiple features, so the rows can sum to more than one. In other words, we assume there are multiple overlapping clusters, and each data point can belong to several clusters simultaneously. IBP corresponds to the latter case.

5. Why we need to perform sampling for inference purposes in the context of Bayesian Nonparametric Models?
   - (Please refer to the class notes and reference papers). In practice, computation of posteriors is not tractable and we need approximate inference methods to compute them. We utilize sampling methods for these approximations.

6. Briefly explain what is Gibbs Distribution.
   - (Please refer to the class notes and reference papers). Gibbs distribution is a probability distribution or probability measure that gives the probability that a system will be in a certain state as a function of that state's energy and the temperature of the system. It is also a frequency distribution of particles in a system.

7. Briefly but concisely explain what MCMC is?
   - (Please refer to the class notes and reference papers). Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution. MCMC constructs a Markov chain that has the desired distribution as its equilibrium distribution, and obtains a sample of the desired distribution by observing the chain after a number of steps. The more steps there are, the more closely the distribution of the sample matches the actual desired distribution.