# Approximate Inference

Sargur Srihari srihari@cedar.buffalo.edu

#### Plan of Discussion

- Need for Approximation
- Types of Approximation

# Central Task in Using Probabilistic Models with latent variables

In unsupervised learning no labelled samples of latent variables

#### 1. Evaluation of posterior distribution p(Z|X)

- Where Z=latent variables, X=observed data variables
  - GMM: Z=latent subclasses  $(z_0,...z_K)$ , X=observed data
  - HMM: Z=latent variables  $(z_0,...z_K)$ , X=observed data

#### 2. Evaluation of expectation of p(X,Z) wrt to p(Z|X)

- E.g., in EM for m.l.e. of parameters of p(X,Z):
  - evaluate expectation of complete-data log-likelihood  $\ln p(X,Z)$  wrt posterior of p(Z/X)

# Need for Approximation for EM

- Often infeasible to evaluate posterior distributions or expectations wrt distributions
  - High dimensionality of latent space
  - Complex and intractable expectations
- In the case of GMMs we get expressions
- Posterior

$$p(Z \mid X, \mu, \Sigma) \alpha \prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_{k} N(x_{n} \mid \mu_{k}, \Sigma_{k}))^{z_{nk}}$$

Expectation

Which are evaluated and maximized using EM

$$E_{Z}\left[\ln p(X,Z\mid\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma})\right] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left\{\ln \boldsymbol{\pi}_{k} + \ln N(x_{n}\mid\boldsymbol{\mu}_{k},\boldsymbol{\Sigma}_{k})\right\}$$

### Difficulties with Posteriors and Expectations

- For continuous variables
  - Required integrations have no closed form solutions
  - Dimensionality of space and integrand prohibit numerical integration
- For discrete variables
  - Summation in marginalization: exponential no. of states

## Example of an Inference Problem

Observed Variables

$$X = \{x_1, ..., x_N\}$$
 N i.i.d. data

Latent Variables and Parameters

$$Z = \{z_1,...,z_N\}$$

- Model for joint distribution p(X,Z) is specified
- Goal is to find approximation for posterior distribution p(Z|X) as well as for p(X)

## Posterior Distribution in Bayesian Methods

Problem of computing posterior is a special case of Variational Inference

$$p(Z \mid X, \alpha) = \frac{p(Z, X \mid \alpha)}{\int_{Z} p(Z, X \mid \alpha)}$$

We cant compute the posterior for many interesting models

Consider the Bayesian mixture of Gaussians,

- 1. Draw  $\mu_{k} \sim N(0, \sigma^{2})$  for k = 1, ... K
- 2. For i = 1...n:
- (a) Draw  $z_i \sim Mult(\pi)$
- (b) Draw  $x_i \sim N(\mu_{z_i}, \sigma^2)$

Suppressing the fixed parameters, the distribution is

$$p(\mu_{1:K}, z_{1:n} \mid x_{1:n}) = \frac{\prod_{k=1}^{K} p(\mu_k) \prod_{i=1}^{n} p(z_i) p(x_i \mid z_i, \mu_{1:K})}{\int \sum_{\mu_{1:K}} \sum_{z_{1:n}} \prod_{k=1}^{K} p(\mu_k) \prod_{i=1}^{n} p(z_i) p(x_i \mid z_i, \mu_{1:K})}$$

- The numerator is easy to compute for any configuration of hidden variables
- The problem is the denominator.

The integral is not easy to compute The summation has  $K^n$  terms which is intractable

- Situation arises in most interesting problems
- Approximate posterior inference is one of the central problems of Bayesian statistics

# Types of Approximations

#### 1. Stochastic

Markov chain Monte Carlo

Have allowed use of Bayesian methods across many domains

Computationally demanding

They can generate exact results

#### 2. Deterministic

- Variational Inference (or Variational Bayes)
- Based on analytical approximations to posterior
  e.g., particular factorization or specific parametric form such as
  - Ġaussian
- Scale well in large applications
- Can never generate exact results