5 Learning Image Attributes

5.1 Gibbs sampling

Given the observations Y (i.e. HOG features of images), we would like to infer the binary matrix Z, which specifies which image contains which attributes, the strengths of images' attributes X, the mixing matrix G, which defines what each attribute looks like, and all of the hyperparameters. We used a collapsed Gibbs sampler [14] to do the inference, but with Metropolis-Hastings steps for sampling new features. At each step, we draw samples from the marginal distribution of the model parameters given the data by successively sampling the conditional distributions of each parameter given all others. We will use the notation $-$ to denote the rest of the variables that are not explicitely conditioned upon in the current state of the Markov chain. The r^{th} row and the c^{th} column of a matrix A will be denoted by $A_{r:}$ and $A_{:c}$, respectively.

First, we note that our likelihood function is

$$P(\mathbf{Y}|\mathbf{X}, \mathbf{G}, \Psi) = \prod_{n=1}^{N} \frac{1}{(2\pi)^{D/2}|\Psi|^{1/2}} \times \exp(-\frac{1}{2}(\mathbf{y}_n - \mathbf{x}_n\mathbf{G})^T\Psi^{-1}(\mathbf{y}_n - \mathbf{x}_n\mathbf{G})). \tag{9}$$

Then we will specify the update formulas for each step in the sampling algorithm, as well as a method for adding new features.

Update for Z and X We first specify the ratio of likelihoods and then integrate out the element x_{nk} to obtain

$$\frac{P(\mathbf{Y}|\mathbf{Z}_{nk} = 1, -)}{P(\mathbf{Y}|\mathbf{Z}_{nk} = 0, -)} = \frac{\int P(\mathbf{Y}|x_{nk}, -)N(x_{nk}; 0, \lambda_k^{-1})dx_{nk}}{P(\mathbf{Y}|x_{nk} = 0, -)} = \sqrt{\frac{\lambda_k}{\lambda}} \exp(\frac{1}{2}\lambda\mu^2), \tag{10}$$

where $\lambda = \Psi_n^{-1} G_{k:}^T G_{k:} + \lambda_k$ and $\mu = \frac{\psi_n^{-1}}{\lambda} G_{k:}^T \hat{E}_n$ with the matrix of residuals $\hat{\mathbf{E}} = \mathbf{Y} - \mathbf{XG}$ evaluated with $x_{nk} = 0$.

From the exchangeability property of the IBP, we can imagine that the n^{th} image was the last to observed, so that the ratio of the priors is

$$\frac{P(Z_{nk} = 1|-)}{P(Z_{nk} = 0|-)} = \frac{m_{-n,k}}{N - 1 - m_{-n,k}}, \tag{11}$$

where $m_{-n,k} = \sum_{s \neq n} z_{s,k}$.

Multiplying equations (10) and (11) gives the expression for the ratio of the posterior probabilities for z_{nk}, which will be used to sample. If z_{nk} is set to 1, we sample $x_{nk}|- \sim N(\mu, \lambda^{-1})$.

Adding new features Let κ_n be the number of columns of $\mathbf{Z}$ which contain 1 only in row n. New features are proposed by sampling κ_n with a Metropolis-Hastings step. We integrate out $\mathbf{G}$ and include $\mathbf{x}$ as part of our proposal. That is, the proposal is $\xi = \{\kappa_n, \mathbf{x}\}$, and a move $\xi \to \xi^*$ is proposed with probability $J(\xi^*|\xi)$. The simplest proposal would be to use the prior on ξ^*

$$J(\xi) = P(\kappa_n|\alpha)P(\mathbf{x}|\kappa_n, \lambda_k) = Poisson(\kappa_n; \gamma)N(\mathbf{x}; 0, \lambda_k^{-1}), \tag{12}$$

where $\gamma = \frac{\alpha}{N-1}$.

The rate constant γ tends to be very small, resulting new features being rarely proposed. This problem can be fixed by introducing two new tunable parameters, π and ω, so the proposal distribution for κ_n becomes

$$J(\kappa_n) = (1 - \pi)Poisson(\kappa_n; \omega\gamma) + \pi\mathbf{1}(\kappa_n = 1) \tag{13}$$

The rest of the updates are straightforward.

Update for G

$$P(\mathbf{g}_d|-) \propto P(\mathbf{y}_d|\mathbf{g}_d, -)P(\mathbf{g}_d) = N(\mathbf{g}_d; \mu_d, \mathbf{\Lambda}), \tag{14}$$

where $\mathbf{\Lambda} = \mathbf{X}^T\Psi^{-1}\mathbf{X} + I$ and $\mu_d = \mathbf{\Lambda}^{-1}\mathbf{X}^T\Psi^{-1}\mathbf{y}_n$.

Update for λ

$$P(\lambda_k|G) \sim Gamma(c + \frac{m_k}{2}, d + \sum_n X_{nk}^2). \tag{15}$$