Assignment 6

1 Mixture of multivariate Bernoullis EM

Consider a unsupervised mixture of multivariate Bernoullis model. For each data example n, we have D binary values x_{nj}, we assume there is an unobserved cluster indicator $y_n \in \{1 \ldots K\}$, and that $P(x_{nj} \sim \text{Ber}(\mu_{y_n,j}))$. In this problem, you will show how to fit this model using EM.

1. Using the definition of arbitrary distributions $q_n(y_n)$, write the evidence lower bound $Q(\theta, \{q_n\})$. Simplify as much as you can.

2. Using Bayes rule, write an expression that can be used to calculate the cluster responsibility $r_{nk} = P(y_n = k|\theta)$

3. Show that the M step for ML estimation of a mixture of multivariate Bernoullis is given by

$$\mu_{kj} = \frac{\sum_n r_{nk} x_{nj}}{\sum_n r_{nk}}$$

4. Show that the M step for MAP estimation of a mixture of multivariate Bernoullis with a $Beta(\alpha, \beta)$ prior is given by

$$\mu_{kj} = \frac{(\sum_n r_{nk} x_{nj}) + \alpha - 1}{(\sum_n r_{nk}) + \alpha + \beta - 2}$$

2 Adding a new feature in least-squares regression

Let $RSS(w) = ||Xw - y||_2^2$ be the residual sum of squares.

a. Show that

$$\frac{\partial}{\partial w_k} RSS(w) = a_k w_k - c_k$$
$$a_k = 2\sum_{i=1}^{n} x_{ik}^2 = 2||x_{:,k}||^2$$
$$c_k = 2\sum_{i=1}^{n} x_{ik}(y_i - w_{-k}^T x_{i,-k}) = 2x_{:,k}^T r_k$$

where $w_{-k} = w$ without component k, $x_{i,-k}$ is x_i without component k, and $r_k = y - w_{-k}^T x_{:,-k}$ is the residual due to using all the features except feature k. Hint: Partition the weights into those involving k and those not involving k.

b. Show that if $\frac{\partial}{\partial w_k} RSS(w) = 0$, then

$$\hat{w}_k = \frac{x_{:,k}^T r_k}{||x_{:,k}||^2}$$

Hence when we sequentially add features, the optimal weight for feature k is computed by computing orthogonally projecting $x_{:,k}$ onto the current residual.