Homework 2

CS839: Probabilistic Graphical Models
Fall 2018, UW-Madison

1 EM for a Mixture of Bernoullis

You need to derive an expectation-maximization (EM) algorithm to cluster black and white images. The inputs $x^{(i)}$ can be thought of as vectors of binary values corresponding to black and white pixel values, and the goal is to cluster the images into groups. You will be using a mixture of Bernoullis model to tackle this problem.

1.1 Mixture of Bernoullis

1. Consider a vector of binary random variables $x \in \{0,1\}^D$. Assume each variable x_d is drawn from a Bernoulli(p_d) distribution, so $P(x_d = 1) = p_d$. Let $p \in (0,1)^D$ be the resulting vector of Bernoulli parameters. Write an expression for $P(x|p)$.

2. Now suppose we have a mixture of K Bernoulli distributions: each vector $x^{(i)}$ is drawn from some vector of bernoulli random variables with parameters $p^(k)$, we will call this Bernoulli($p^{(k)}$). Let $\{p^{(1)}, p^{(2)}, \ldots, p^{(K)} = \mathbf{p}\}$. Assume a distribution $\pi(k)$ over the selection of which set of Bernoulli parameters $p^{(k)}$ is chosen. Write the expression for $P(x^{(i)}|\mathbf{p}, \pi)$.

3. Finally, suppose we hae inpute $X = \{x^{(i)}\}_{i=1\ldots n}$. Using the above, write an expression for the log likelihod of the data X, $\log P(X|\pi, \mathbf{p})$.

1.2 Expectation Step

1. Now, we introduce the latent variables for the EM algorithm. Let $z^{(i) \in \{0,1\}^K}$ be an indicator vector, such that $z^(i)_k = 1$ if $x^{(i)}$ was drawn from a Bernoulli ($p^{(k)}$) and 0 otherwise. Let $Z = \{z^{(i)}_{i=1,\ldots,n}\}$. What is $P(z^{(i)}|\pi)$? What is $P(z^{(i)}|\mathbf{p}, \pi)$?

2. Using the above two quantities, derive the likelihood of the data and the latent variables $P(Z, X|\pi, \mathbf{p})$.

3. Let $\eta(z_k^{(i)}) = E[z_k^{(i)}|x^{(i), \pi, \mathbf{p}}]$. Show that:

$$\eta(z_k^{(i)}) = \frac{\pi_k \prod_{d=1}^{D} (p_d^{(k)})^{x_d^{(i)}} (1 - p_d^{(k)})^{1 - x_d^{(i)}}}{\sum_j \pi_j \prod_{d=1}^{D} (p_d^{(j)})^{x_d^{(i)}} (1 - p_d^{(j)})^{1 - x_d^{(i)}}} \tag{1}$$

   Let $\tilde{\mathbf{p}}, \tilde{\pi}$ be the new parameters that we'd like to maximize, so $\mathbf{p}$, π are from the previous iteration. Use this to derive the following final expression for E-step in the expectation-maximization algorithm:

$$E[\log P(Z, X|\tilde{\mathbf{p}}, \tilde{\pi})|X, \mathbf{p}, \pi] = \sum_{i=1}^{N} \sum_{k=1}^{K} \eta(z_k^{(i)}) \left[\log \tilde{\pi}_k + \sum_{d=1}^{D} \left(x_d^{(i)} \log \tilde{p}_d^{(k)} + (1 - x_d^{(i)}) \log(1 - \tilde{p}_d^{(k)}) \right) \right] \tag{2}$$

1.3 Maximization Step

1. We need to maximize the above expression with respect to $\tilde{\mathbf{p}}, \tilde{\pi}$. First, show that the value of $\tilde{\mathbf{p}}$ that maximizes the E step is:

$$\tilde{p}^{(k)} = \frac{\sum_{i=1}^{N} \eta(z_k^{(i)}) x^{(i)}}{N_k} \tag{3}$$

   where $N_k = \sum_{i=1}^{N} \eta(z_k^{(i)})$.

2. Show that the value of $\tilde{\pi}$ that maximizes the E-step is:

$$\tilde{pi}_k = \frac{N_k}{\sum_{k'} N_{k'}} \tag{4}$$

   The exponential families notation may be useful. Alternatively, you can use Lagrance multipliers.

2 Estimation of Precision Matrix

Let's learn some structures. We denote the precision matrix as $P = \Sigma^{-1}$ where Σ is the covariance matrix of Gaussian data.

1. Many relevant pieces of literature state that "it is well known that the zero coefficients in P correspond to conditional independence of corresponding variables", but they barely show why. Let's try to do it here: show that the coefficient in precision matrix $P_{i,j}$ encodes partial correlation relationship of corresponding variables $p_{i,j}$.

2. Let's take a step further about partial correlation: show that partial correlation of two variables $p_{i,j}$ is proportional to linear dependency coefficient $\phi_{i,j}$ where $X_j = X_i \phi_{i,j}$.

3. Show that estimating the graph is effectively optimizing the next equation:

$$\arg\min_{P} - \log\det P + \mathbf{tr}(\tilde{\Sigma} P) \tag{5}$$

   where $\tilde{\Sigma}$ is the empirical covariance matrix.