NAME - PRASHAM D. SHETH

UNI - pds2136

Problem 1 :—

→ We have labeled set of data
$(y_1, x_1), \ldots, (y_n, x_n)$

$y \in \{0, 1\}$

x is D-dimensional vector

We predict y_0 for new x_0 as

$$y_0 = \arg\max_y \; p(y_0 = y \mid \pi) \prod_{d=1}^{D} p(x_{0,d} \mid \lambda_{y,d})$$

$$p(y_0 = y \mid \pi) = \text{Bernoulli}(y \mid \pi)$$

∴ Data: $y_i \overset{iid}{\sim} \text{Bern}(\pi)$

$x_{i,d} \mid y_i \sim \text{Pois}(\lambda_{y_i,d}) \qquad d = 1 \ldots D$

Prior : $\lambda_{y,d} \overset{iid}{\sim} \text{Gamma}(2, 1)$

$$\hat{\pi}, \hat{\lambda}_{0,1:D}, \hat{\lambda}_{1,1:D} = \arg\max_{\hat{\pi}, \hat{\lambda}_{0,1:D}, \hat{\lambda}_{1,1:D}} [L]$$

$$L = \Big[\sum_{i=1}^{n} \ln p(y_i \mid \pi) + \sum_{d=1}^{D} \Big(\ln p(\lambda_{0,d}) + \ln p(\lambda_{1,d}) + \sum_{i=1}^{n} \ln p(x_{i,d} \mid \lambda_{y_i,d}) \Big)$$

(a) for $\hat{\pi}$, By using $p(y_i|\pi) = \pi^{y_i}(1-\pi)^{1-y_i}$

using first order characteristic,

$$\frac{\partial L}{\partial \hat{\pi}} = \frac{\partial}{\partial \hat{\pi}}\left[\sum_{i=1}^{n}\left(y_i \log \pi + (1-y_i)\log(1-\pi)\right) + c\right]$$

(here c represents the 2nd term in the objective which is a constant w.r.t π)

$$\therefore \frac{\partial L}{\partial \hat{\pi}} = \frac{\sum_{i=1}^{n} y_i}{\pi} + (-1)\frac{\sum_{i=1}^{n}(1-y_i)}{1-\pi} = 0$$

$$\therefore \frac{\sum_{i=1}^{N} y_i}{\pi} - \frac{(n - \sum_{i=1}^{n} y_i)}{1-\pi} = 0$$

$$\therefore \frac{\sum_{i=1}^{N} y_i}{\pi} - \frac{n}{1-\pi} + \frac{\sum_{i=1}^{N} y_i}{1-\pi} = 0$$

$$\therefore \sum_{i=1}^{N} y_i \left[\frac{1}{\pi} + \frac{1}{1-\pi}\right] = \frac{n}{1-\pi}$$

$$\therefore \frac{\sum_{i=1}^{n} y_i}{\pi(1-\pi)} = \frac{n}{1-\pi}$$

$$\therefore \boxed{\hat{\pi} = \sum_{i=1}^{N} y_i \Big/ N}$$

(b) $\lambda_{y,d} \sim \text{Gamma}(2,1)$

$$p(\lambda_{y,d}) = \frac{(1)^2}{\Gamma(2)} \cdot (\lambda_{y,d})^1 \cdot e^{-\lambda_{y,d}}$$

as $\Gamma(2) = 1$

$$p(\lambda_{y,d}) = (\lambda_{y,d}) \cdot e^{-\lambda_{y,d}}$$

$x_{i,d} | y_i \sim \text{Pois}(\lambda_{y_i,d})$

$$p(x_{i,d} | y_i) = \frac{(\lambda_{y_i,d})^{x_{i,d}} \cdot e^{-\lambda_{y_i,d}}}{(x_{i,d})!}$$

We can write L as

$$L = c' + \sum_{d=1}^{D} \Big(\ln p(\lambda_{0,d}) + \ln p(\lambda_{1,d}) + \sum_{i=1}^{n} \ln p(x_{i,d} | \lambda_{y_i,d}) \Big)$$

c' is first term of L which is a constant w.r.t $\lambda_{0,1:D}$ & $\lambda_{1,1:D}$

Also, $$\sum_{i=1}^{N} \ln p(x_{i,d} | \lambda_{y_i,d}) = \sum_{i=1}^{N} \Big(y_i \cdot \ln p(x_{i,d} | \lambda_{1,d}) + (1-y_i) \cdot \ln p(x_{i,d} | \lambda_{0,d}) \Big)$$

(here $y_i = 1$ can be used as indicator for $\lambda_{1,d}$

and $(1-y_i) = 1 \Rightarrow y_i = 0$ can be used as indicator for $\lambda_{0,d}$.

$$\therefore\ L = c' + \sum_{d=1}^{D} \Big(\ln p(\lambda_{0,d}) + \ln p(\lambda_{1,d})$$
$$+ \sum_{i=1}^{n} \Big(y_i \cdot \ln p(x_{i,d} \mid \lambda_{1,d}) + (1-y_i) \cdot \ln p(x_{i,d} \mid \lambda_{0,d}) \Big)\Big)$$

differentiating L w.r.t $\lambda_{0,d}$, (leaving d arbitrary)

$$\frac{\partial L}{\partial \hat{\lambda}_{0,d}} = \frac{d}{d\hat{\lambda}_{0,d}} \Big[c' + \sum_{d=1}^{D} \ln p(\lambda_{0,d}) + \ln(\lambda_{1,d})$$
$$+ \sum_{i=1}^{n} \Big(y_i \cdot \ln p(x_{i,d} \mid \lambda_{1,d}) + (1-y_i) \ln p(x_{i,d} \mid \lambda_{0,d}) \Big)\Big]$$

$$\ln p(\lambda_{0,d}) = \log(\lambda_{0,d}) - \lambda_{0,d}$$

$$\ln p(\lambda_{1,d}) = \log(\lambda_{1,d}) - \lambda_{1,d}$$

$$\ln p(x_{i,d} \mid \lambda_{1,d}) = x_{i,d} \log(\lambda_{1,d}) - \lambda_{1,d} - \log(x_{i,d})!$$

$$\ln p(x_{i,d} \mid \lambda_{0,d}) = x_{i,d} \log(\lambda_{0,d}) - \lambda_{0,d} - \log(x_{i,d})!$$

$$\therefore\ \frac{\partial L}{\partial \hat{\lambda}_{0,d}} = 0$$

$$\therefore\ 0 = -1 + \frac{1}{\lambda_{0,d}} + \sum_{i=1}^{n} (1-y_i) \cdot \left[\frac{x_{i,d}}{\lambda_{0,d}} - 1 \right]$$

$$\therefore\ \hat{\lambda}_{0,d} = \frac{\sum_{i=1}^{n} (1-y_i) \cdot x_{i,d} + 1}{\sum_{i=1}^{n} (1-y_i) + 1}$$

Similarly, for $\hat{\lambda}_{1,d}$

$$\frac{\partial L}{\partial \hat{\lambda}_{1,d}} = -1 + \frac{1}{\lambda_{1,d}} + \sum_{i=1}^{n} y_i \left[\frac{x_{i,d}}{\lambda_{1,d}} - 1 \right]$$

$$\therefore \quad \hat{\lambda}_{1,d} = \frac{\sum_{i=1}^{n} y_i \cdot x_{i,d} + 1}{\sum_{i=1}^{n} y_i + 1}.$$

$\therefore$ from the above forms of $\hat{\lambda}_{1,d}$ and $\hat{\lambda}_{0,d}$

$$\boxed{\hat{\lambda}_{y,d} = y \left[\frac{\sum_{i=1}^{N} y_i \cdot x_{i,d} + 1}{\sum_{i=1}^{N} y_i + 1} \right] + (1-y) \left[\frac{\sum_{i=1}^{N} (1-y_i) x_{i,d} + 1}{\sum_{i=1}^{N} (1-y_i) + 1} \right]}$$