Name : Abhas Kumar
Roll no : 20111001
MTECH, CSE, IITK

BONUS ASSIGNMENT

Q.1 (PCA using an alternate way)

If the given matrix is of form $M = \frac{1}{N}(XX^T)$ with an eigenvector $v \in \mathbb{R}^N$, Then we have

$$Mv = \lambda v$$

i.e $\frac{1}{N}(XX^T)v = \lambda v \Rightarrow X\left\{\frac{1}{N}(XX^T)\right\} = X(\lambda v)$

$$\Rightarrow \frac{1}{N}(X^TX)(X^Tv) = \lambda(X^Tv)$$

Let $u = X^Tv$

then $\frac{1}{N}(X^TX)u = \lambda u$

↳ which is in the form of $Au = \lambda u$, where A is a Matrix, hence u is eigenvector for matrix ~~of~~ A.

$\Rightarrow u$ is eigenvector of $\frac{1}{N}(X^TX)$.

$\Rightarrow X^Tv$ is an eigenvector of $\frac{1}{N}(X^TX)$

* for the case $D > N$, complexity for decomposition of $\frac{1}{N}(XX^T)$ will be $O(KN^2)$ and complexity of multi.. matrix is $O(KND)$ where K is the number of eigenvectors to be computed.
* So, for the normal case ~~complexity~~ complexity to compute K eigenvectors for Matrix $S = \frac{1}{N}(X^TX)$ is $O(KD^2)$, and for the given case (i.e $D > N$) it will be $O(KN^2) + O(KND) \simeq O(KND)$ which better than the normal case. $O(KD^2)$ since $N < D$.

Q.2 (EM for Poisson Mixture Model)

We have

$$p(K, z \mid \lambda, \pi) = \prod_{n=1}^{N} \prod_{l=1}^{L} \left[p(z_n = l) \prod_{m=1}^{M} \text{Poisson}(K_{n,m} \mid \lambda_l) \right]^{\mathbb{I}[z_n = l]}$$

where Poisson has the form $p(K \mid \lambda) = \frac{1}{e^\lambda} \frac{\lambda^k}{K!}$ with parameter λ can count λk, L is the number of clusters

$K_{n,m}$ is number of hits to the web server n in minutes m.

If latent the latent variables z_n are sampled from multinoulli distribution, and K_{nm} are sampled from Poisson distribution, with $\lambda = \{\lambda_1, \lambda_2, ---, \lambda_L\}$ are parameters of the Poisson distribution i.e

$$K_{nm} \sim \text{Poisson}(K_{nm} \mid \lambda_l)$$

Let $\theta = \{\pi_l, \lambda_l\}_{l=1}^{L}$ be the parameters to be learned, then complete data log likelihood for the model can be given as.

$$\text{CLL} = \underset{\theta}{\text{argmax}} \left(\log p(K, z \mid \theta) \right) = \log \prod_{n=1}^{N} \prod_{l=1}^{L} \left[p(z_n = l) \prod_{m=1}^{M} \text{Poisson}(K_{n,m} \mid \lambda_l) \right]^{\mathbb{I}[z_n = l]}$$

$$\therefore \text{CLL} = \underset{\theta}{\text{argmax}} \left\{ \sum_{n=1}^{N} \sum_{l=1}^{L} E[z_{nl}] \left[\log \pi_l + \sum_{m=1}^{M} \log \left[\text{Poisson}(K_{nm} \mid \lambda_l) \right] \right] \right\} \quad (1)$$

where the $\text{Poisson}(K_{nm} \mid \lambda_l) = \dfrac{\lambda_l^{K_{nm}} e^{-\lambda_l}}{K_{nm}!}$ using $\frac{1}{e^\lambda} \times \frac{\lambda^K}{K!}$

also z_n is L-dimensional vector

where $z_{nl} = 1$ means z_n belongs to cluster l.

and $E[z_{nl}]$ is the Expectation of z_{nl}

Algorithm to estimate z_{nl}.

1. Initialise $\theta = \{\pi_l, \lambda_l\}_{l=1}^{L}$.

2. compute expectation of each z_n

$$E[z_{nl}] = \frac{\pi_l^{(t-1)} \prod_{m=1}^{M} \text{poisson}(k_{nm} \mid \lambda_l)}{\sum_{l=1}^{L} \pi_l^{(t-1)} \prod_{m=1}^{M} \text{poisson}(k_{nm} \mid \lambda_l)}$$

3. To estimate θ i.e π_l, λ_l

$$\pi_l = \frac{N_l}{N} \quad \text{where } N_l = \sum_{n=1}^{N} E[z_{nl}]$$

also

$$\frac{\partial CLL}{\partial \lambda_l} = \frac{d}{d\lambda_l}\left(\sum_n \sum_l E[z_{nl}]\left[\log \pi_l + \sum_m \log\left(\frac{\lambda^{k_{nm}} e^{-\lambda_l}}{k_{nm}!}\right)\right]\right)$$

$$= \frac{d}{d\lambda_l}\left(\sum_n \sum_l E[z_{nl}]\left[\log \pi_l + \sum_m k_{nm} \log \lambda_l - \lambda_l - \log(k_{nm}!)\right]\right)$$

$$\Rightarrow \sum_n \sum_l E[z_{nl}]\left[\sum_m \left(\frac{k_{nm}}{\lambda_l} - 1\right)\right] = 0$$

$$\Rightarrow \sum_n \sum_l \sum_m E[z_{nl}] \frac{k_{nm}}{\lambda_l} = \sum_n \sum_l M \times E[z_{nl}]$$

$$\Rightarrow \lambda_l = \frac{\sum_n \sum_l \sum_m E[z_{nl}] \cdot k_{nm}}{\sum_n \sum_l M E[z_{nl}]}$$

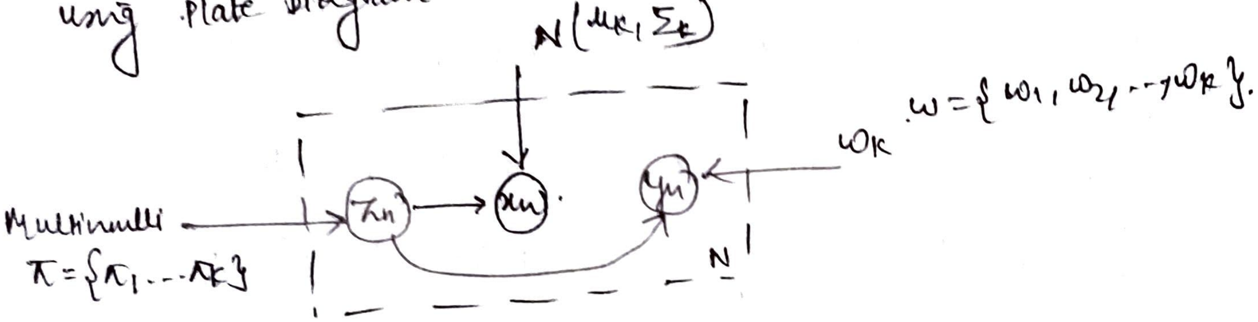
4. Repeat step 2 and 3 until not converged.

Q.3 (A Latent variable Model for Regression).

Part (1) ~~let the~~

a. The proposed model can be thought of combination of many different linear curves as this model is first clustering the data on different linear curves and then the prediction are made for y. In contrast a standard linear model will only work for those situation where we have to regress a linear curve only.

Input-output relation can be shown for the proposed model using Plate Diagram.



b. let $\theta = \{\pi_k, \mu_k, \Sigma_k, w_k\}_{k=1}^{K}$

$$P(z_n, x_n, y_n, y_n \mid \theta) = \pi_k \times N(\mu_k, \Sigma_k) \times N(w_k^T x_n, \beta^{-1})$$

Complete data log likelihood can be given as.

$$CLL = \log \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k \times N(\mu_k; \Sigma_k) \times N(w_k^T x_n, \beta^{-1})\right]^{I(z_n=k)}$$

$$CLL = \sum_{n=1}^{N} \sum_{k=1}^{K} I(z_n=k)\left[\log N(\mu_k, \Sigma_k) + \log \pi_k + \log N(w_k^T x_n, \beta^{-1})\right]$$

Expectation of CLL with respect to distribution $P(z_n \mid y_n, x_n, \theta)$

$$= \sum_{n} \sum_{k} E[I(z_n=k)]\left[\log N(\mu_k, \Sigma_k) + \log \pi_k + \log N(w_k^T x_n, \beta^{-1})\right]$$

$$\propto \pi_k \times N(\mu_k, \Sigma_k) \times N(w_k^T x_n, \beta^{-1})$$

$$\hat{\pi}_k = \frac{1}{N}\sum_{n=1}^{N} E[I(z_n=k)]$$

$$\hat{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N} E[I(z_n=k)]\, x_n$$

$$\hat{\Sigma}_k = \frac{1}{N_k}\sum_{n=1}^{N} E[I(z_n=k)]\,(x_n-\hat{\mu}_k)(x_n-\hat{\mu}_k)^T$$

$$\hat{w}_k = \left(\sum_{n=1}^{N} \epsilon_n x_n x_n^T\right)^{-1}\left(\sum_{n=1}^{N} \epsilon_n y_n x_n\right)$$

where $\epsilon_n = E[I(z_n=k)]$.

EM algo

Step1: initialise $\theta = \{\pi_k, \mu_k, \Sigma_k, w_k\}_{k=1}^{K}$ at $t=0$

Step2: compute posterior of latent variables given $\theta^{(t-1)}$.

$$\frac{\pi_k^{(t-1)} N(\mu_k, \Sigma_k) \times N(w_k^T x_n, \beta^{-1})}{\sum_{l=1}^{K} \pi_l^{(t-1)} N(\mu_k, \Sigma_l) \times N(w_l^T x_n, \beta^{-1})}$$

Step 3: Maximise the expected CLL w.r.t to θ keeping $z_n^{(t)}$ constant

Step 4: $t = t+1$

Step 5: Repeat step 2 and 3+4 until converged.

* update of w_k makes sense as we are getting expected least square solution based on posterior, ~~if only~~ y_n, taken x_n distribution into account.

Now, $\pi_k = 1/K \; \forall k$.

ALT-OPT

Step1: initialise $\theta = \{\mu_k; \Sigma_k, w_k\}$. at $t=0$

Step2:

$$z_n^{(t)} = \arg\max_{k=1,2,\dots K} P(z_n = k \mid \theta^{t-1}, x_n, y_n)$$

Step 3: Estimate MLE for CLL taking $z_n^{(t)}$ as fixed.

Step4: $t = t+1$.

Step5: if not converged go to step 2.

Part 2 :

Let $\theta = \{\eta_k, w_k\}_{k=1}^{K}$ and $\pi_k(x_n) = \dfrac{\exp(\eta_k^T x_n)}{\sum_{l=1}^{K} \exp(\eta_l^T x_n)}$

$$P(z_n, y_n \mid x_n, \theta)$$
$$= P(z_n = k) \times P(y_n \mid z_n, x_n, \theta)$$
$$= \pi_k(x_n) \times N(w_k^T x_n, \beta^{-1})$$

$$CLL = \log \prod_{n=1}^{N} \prod_{k=1}^{K} \left[P(z_n, y_n \mid x_n, \theta) \right]^{I(z_n = k)}$$

$$E(CLL) = \sum_{n=1}^{N} \sum_{k=1}^{K} E\left[I(z_n = k)\right] \times \left[\log \pi_k(x_n) + \log N(w_k^T x_n, \beta^{-1}) \right]$$

$E(I(z_n = k))$ w.r.t. $p(z_n \mid y_n, x_n, \theta) \propto \pi_k(x_n) \times N(w_k^T x_n, \beta^{-1})$.

~~Part 2~~ where x_n is not modelled, z_n is multinoulli$(\pi_1(x_n), \ldots \pi_K(x_n))$

where $\pi_k(x_n)$ is defined as softmax function $\dfrac{\exp(\eta_k^T x_n)}{\sum_{l=1}^{K} \exp(\eta_l^T x_n)}$

Given, parameter $\theta = \{\eta_k, w_k\}_{k=1}^{K}$

EM algorithm

step1: initialise $\{\pi_k, w_k\}_{k=1}^{K} = \theta$

Step 2: Compute
$$P(z_n = k \mid y_n, \theta) = \frac{\pi_k N(w_k^T x_n, \beta^{-1})}{\sum_{l=1}^{K} \pi_l N(w_l^T x_n, \beta^{-1})}$$

$$\Rightarrow p(z_n = k \mid y_n, \theta) = \frac{\pi_k \exp\left(-\frac{\beta}{2}(y_n - w_k^T x_n)^2\right)}{\sum_{l=1}^{K} \pi_l \left(\exp\left(-\beta/2\,(y_n - w_l^T x_n)^2\right)\right)}$$

$\cancel{\eta_k = \sum_{n=1}^{N} E[z_{nk}]}$, $\cancel{w_k = (X_k^T X_k)^{-1} X_k^T y_k}$, $\cancel{\pi_k = N_k}/N$.

Step3: Maximise the Expectation i.e $E(CLL)$

step4: $t = t+1$, and repeat 2 and 3 until converged.

* There is no closed form solution for η_k in this model in M step, iterative methods needs to be used.