Bonus Assignment

CS771A - 2020 Autumn

problem 1 :

Given $X_{N \times D}$ u eigen vectors of $\frac{1}{N} X^T X$; v eigen vectors of $\frac{1}{N} X X^T$

we have eigen vectors v of $\frac{1}{N} X X^T$

$$\Rightarrow \left[\frac{1}{N} X X^T\right] v = \lambda v$$

$$\Rightarrow X^T \left[\frac{1}{N} X X^T\right] v = X^T \lambda v \qquad \{\text{multiplying by } X^T\}$$

$$\Rightarrow \frac{1}{N}\left[X^T X\right]\left[X^T v\right] = \lambda \left[X^T v\right]$$

$\therefore$ $X^T v$ is eigen vectors of $\frac{1}{N} X^T X$

$\therefore$ $X^T v = u$

Advantage : Since computation of eigen vectors of XX^T i.e $v \in \mathbb{R}^N$ will be cheaper as $N < D$. It is cheaper as it required to eigen decomposition XX^T i.e $N \times N$ matrix than $X^T X$ i.e $D \times D$ matrix

And also we can obtain u from v.

Bonus Assignment

CS771 - 2020 Autumn

problem 2:

Given $p(k|\lambda) = \frac{1}{e^{\lambda}} \frac{\lambda^k}{k!}$. λ - is $\frac{1}{N}\sum_n k_n$ (mean of counts)

N - observations

L - no. of clusters

for latent model

$P(z_n=l)$ - mixing probability

$\theta = \{\pi_l, \lambda_l\}_{l=1}^{L}$; $\sum_{l=1}^{L} \pi_l = 1$

$$\text{poision}(k_{nm} | \lambda_l) = \frac{(\lambda_l)^{k_{n,m}} e^{-\lambda_l}}{k_{nm}!}$$

$\forall n \in [1..N]$

$\forall m \in [1..m]$

$\forall l \in \{1..L]$

likelihood

user transaction x_i from X in cluster l $x_i \in \mathbb{R}^M$

is $p(x_i|\lambda_l) = \prod_{j=1}^{M} \frac{(\lambda_l)^{x_{ij}} e^{-\lambda_l}}{x_{ij}!}$

hitting the server ie $(x_{i1}, x_{i2} .. x_{im})$

x_{ij} - time spent by user on server $x \cdot K_{nm}$

$\therefore P(X = x_i | \lambda) = \sum_{l=1}^{L} P(z_n = l | \lambda) p(X = x_i | z_n = l, \lambda_l)$

$= \sum_{l=1}^{L} \pi_l P(X = x_i | z_n = l, \lambda_l)$

$\therefore P(x_i | \lambda_l) = \sum_{l=1}^{L} \pi_l \left(\prod_{j=1}^{M} \frac{(\lambda_l)^{x_{ij}} e^{-\lambda_l}}{x_{ij}!} \right)$

likelihood $L(X|\theta) = \prod_{i=1}^{N} \sum_{l=1}^{L} \pi_l \cdot \prod_{j=1}^{M} \frac{(\lambda_l)^{x_{ij}} e^{-\lambda_l}}{x_{ij}!}$

log likelihood $= \sum_{i=1}^{N} \log \sum_{l=1}^{L} \pi_l \prod_{j=1}^{M} \frac{(\lambda_l)^{x_{ij}} e^{-\lambda_l}}{\ldots}$

Maximizing log likelihood

$$\frac{\partial l(x|\theta)}{\partial \lambda_l} = \sum_{i=1}^{N} \frac{\pi_l}{\sum_{l=1}^{L} \pi_l f(x_i|\lambda_l)} \cdot \left[\frac{x_i [\lambda_l]^{x_i - 1} e^{-\lambda_l} - \lambda_l^{x_i} e^{-\lambda_l}}{x_i !} \right]$$

where

$$f(x_i, \lambda_i) = \frac{(\lambda_l)^{x_i} e^{-\lambda_l}}{x_i !}$$

$$= \sum_{i=1}^{N} \frac{\pi_l}{\sum_{l=1}^{L} \pi_l f(x_i|\lambda_l)} \left[\frac{x_i}{\lambda_l} f(x_i|\lambda) - f(x_i|\lambda_l) \right]$$

$$= \sum_{i=1}^{N} \frac{x_i}{\lambda_l} f(l|x_i, \lambda_l) - f(l|x_i \lambda_l) \Big]$$

$$\Rightarrow \lambda_{l+1} = \lambda_l + S_\lambda(\theta_l) \frac{\partial l(X, \theta)}{N \partial \lambda}$$

$$S_\lambda(\theta) = \mathrm{diag}\left(\frac{N\lambda_1}{\sum f(l|x_i, \theta)} \cdots \frac{N\lambda_L}{\sum f(l|x_i, \theta)} \right)$$

$$\pi_{l+1} = \pi_l + S_\pi(\theta_l) \frac{\partial l(X, \theta)}{N \partial \pi}$$

$$S_\pi(\theta_l) = \mathrm{diag}(\pi_1, \pi_2 \cdots \pi_L) - \pi \pi^T$$

Bonus Assignment

CS771A - 2020 Autumn

Problem 3:

Given $\{(x_n, y_n)\}_{n=1}^{N}$ $\quad x_n \in \mathbb{R}^D$, $y_n \in \mathbb{R}$

N training Examples

The Generative Model: such that

$$z_n \sim \text{multinoulli}(\pi_1, \pi_2 \ldots \pi_K)$$
$$x_n \sim \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})$$
$$y_n \sim \mathcal{N}(w_{z_n}^T x_n, \beta^{-1})$$

Part 1

(1) In the linear regression model that we have learnt, it will gives only one decision boundary for all the datapoints. (shown in fig 1). when we introduced z_n with each input (z_n a latent variable) the model learns multiple linear regressions i.e multiple boundaries. shown in fig 2.

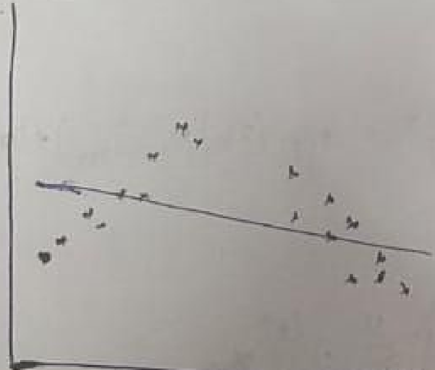

fig 1: standard linear Regression

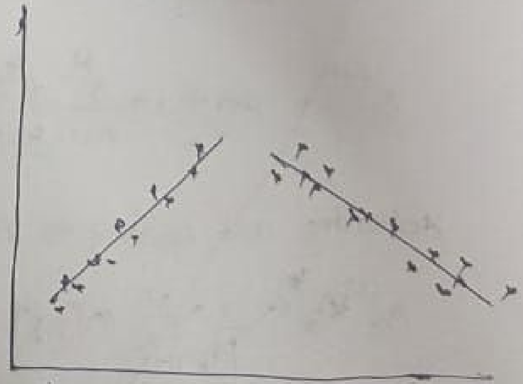

fig 2: latent variable linear Regression.

i.e Trying to divide the dataset into different classes & fit linear regression to each class.

$z_n \sim$ multin .

CLL can be written as

$$\sum_{n=1}^{N} \log p(y_n, z_n, x_n|\theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left[\log \pi_k + \log \mathcal{N}(x_n|\mu_{z_n}, \Sigma_{z_n}) + \log \mathcal{N}(y_n|w_{z_n}^T x_n, \beta^{-1})\right]$$

Now EM Algorithm

① Initialize $\theta = \{\pi_k, \mu_k, \Sigma_k, w_k\}_{k=1}^{K}$ as θ^0 set $t=1$

② E step

$$P(z_n=k|y_n, x_n, \theta) = \gamma_{nk}^{(t)} = P(Z_n=k|x_n, y_n\theta)$$

$$= \frac{P(Z_n=k|\theta)\, P(x_n|Z_n=k,\theta)\, P(y_n|Z_n,\theta,x_n)}{\sum_k P(Z_n=k|\theta)\, P(x_n|Z_n=k,\theta)\, P(y_n|Z_n,\theta,x_n)}$$

$$= \frac{\pi_k \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})\, \mathcal{N}(w_{z_n}^T x_n, \beta^{-1})}{\sum_k \pi_k \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})\, \mathcal{N}(y_n|w_{z_n}^T x_n, \beta^{-1})}$$

③ M step - Updating θ

$$\theta^{t+1} = \arg\max_{\theta} \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{E}[z_{nk}]\left[\log \pi_k + \log \mathcal{N}(\mu_{z_n}, \Sigma_n) + \log \mathcal{N}(w_{z_n}^T x_n, \beta^{-1})\right]$$

deriving & equating to zero we get

$$\mu_k^{(t)} = \frac{1}{N_k}\sum_{n=1}^{N} \gamma_{nk}^{(t)} x_n$$

$$\Sigma_k^{(t)} = \frac{1}{N_k}\sum_{n=1}^{N} \gamma_{nk}^{(t)} (x_n - \mu_k^{(t)})(x_n - \mu_k^{(t)})^T$$

$$\pi_k = \frac{N_k}{N}$$

Now for w_k

$$w_k = \left(\sum_{n=1}^{N}\sum_{k=1}^{k} \gamma_{nk} x_n x_n^T\right)^{-1}\left(\sum_{n=1}^{N}\sum_{k=1}^{k} \gamma_{nk} y_n x_n\right)$$

$t = t+1$

Repeat till convergence

The w_k udate makes intitute that its considering the weighted x_n's based on the cluster it belong to update w_k.

And it should be senario.

Now - ALT OPT when $\pi_k = 1/k$

① initilize θ as θ^0 and set $t=0$

② find z_n

$$z_n^{(t+1)} = \arg\max_k \frac{\mathcal{N}(\mu_{z_n}, \Sigma_{z_n})\,\mathcal{N}(w_k^T x_n, \beta^{-1})}{\sum_{\ell=1}^{k} \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})\,\mathcal{N}(w_{z_n}^T x_n, \beta^{-1})}$$

This update is ignoring no. of points in each cluster wrt w_k and giving them equal weights

③ θ_{z_n} update

$$\theta_{MLE}^{(t+1)} = \arg\max_{\theta} \sum_{n=1}^{N}\sum_{k=1}^{k} z_{nk}^{(t+1)}\left[\log \mathcal{N}(\mu_{z_n}, \Sigma_{z_n}) + \log \mathcal{N}(w_{z_n}^T x_n, \beta^{-1})\right]$$

$$\Rightarrow \mu_k^{(t+1)} = \frac{1}{N_k}\sum_{n=1}^{N} z_{nk}^{(t+1)} x_n$$

$$\Sigma_k^{(t+1)} = \frac{1}{N_k}\sum_{n=1}^{N} z_{nk}(x_n - \mu_k^{(t+1)})(x_n - \mu_k^{(t+1)})^T$$

part (2) Given x_n's

$z_n \sim \text{multinoulli}(\pi_1(x_n), \ldots \pi_K(x_n))$ where

$$\pi_k(x_n) = \frac{\exp(\eta_k^T x_n)}{\sum_{l=1}^{L} \exp(\eta_l^T x_n)}$$

EM Algo

(1) Initialize $\theta = \{\eta_k, w_k\}_{k=1}^{K}$

(2) E step

$$P(z_n = k \mid x_n, y_n, \theta) = \gamma_{nk}^{(t)}$$
$$= \frac{\pi_k(x_n) \mathcal{N}(w_k^T x_n, \beta^{-1})}{\sum_{k=1}^{K} \pi_k(x_n) \mathcal{N}(w_k^T x_n, \beta^{-1})}$$

(3) M step

$$\hat{\theta} = \arg\max_{\theta} \sum_{n=1}^{N} \mathbb{E}[z_{nk}] \left[\log \pi_k(x_n) + \log \mathcal{N}(w_k^T x_n, \beta^{-1})\right]$$

derivating & equating to zero wrt η_l

$$\frac{\partial}{\partial \eta_l} \sum_{n=1}^{N} \mathbb{E}[z_{nk}] \left[\log \pi_k(x_n) + \log \mathcal{N}(w_k^T x_n, \beta^{-1})\right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} \mathbb{E}(z_{nk}) \frac{\partial}{\partial \eta_l} \left(\eta_l^T x_n - \log \sum_{l=1}^{L} \exp(\eta_l^T x_n)\right) = 0$$

$$= \sum_{n=1}^{N} \mathbb{E}[z_{nk}] \left(x_n - \frac{\exp(\eta_l^T x_n) \cdot x_n}{\sum_{l=1}^{L} \exp(\eta_l^T x_n)}\right) = 0$$ No closed form for η_l

slly $\frac{\partial}{\partial w_k} = 0$

$$w_k = \left(\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} x_n x_n^T\right)^{-1} \left(\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} y_n x_n\right)$$