Ahmed Bilal

ABilal19

a) $p(x_2) = N(\mu_1, \Sigma_{11}) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}\right)$

b) $p(x_2) = N(\mu_2, \Sigma_{22}) = N(5, 14)$

c) we know that $p(x_1 | x_2) = N(\mu_{1|2}, \Sigma_{1|2})$ * saw soln

$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \frac{1}{14}\begin{bmatrix} 5 \\ 11 \end{bmatrix}(x_2 - 5)$

while $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix} - \frac{1}{14}\begin{bmatrix} 5 \\ 11 \end{bmatrix}\begin{bmatrix} 5 & 11 \end{bmatrix}$

* saw solution to check answer $= \begin{bmatrix} 59/14 & 57/14 \\ 57/14 & 61/14 \end{bmatrix}$

d) we know that

$p(x_2 | x_1) = N(\mu_{2|1}, \Sigma_{2|1})$

$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1) = 5 + \begin{bmatrix} 5 & 11 \end{bmatrix}\begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}^{-1}(x_1 - \mu_1)$

* checked soln: $= 5 + \begin{bmatrix} -\frac{23}{14} & \frac{13}{7} \end{bmatrix} x_1 - 0$

$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 14 - \begin{bmatrix} 5 & 11 \end{bmatrix}\begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}^{-1}\begin{bmatrix} 5 \\ 11 \end{bmatrix}$

$= \frac{25}{14}$

2a) ☆ saw soln:

$P(y=1 \mid x, \theta) = \sigma(\theta^T x)$

with $nll(\theta) = -\sum_i y_i \log \sigma(\theta^T x) + (1-y_i)\log(1-\sigma(\theta^T x)) + \frac{\lambda}{2}\|\theta\|_2^2$

$\nabla \theta' = \sum_i y_i (1-\sigma(\theta^T x))\, x - (1-y_i)\sigma(\theta^T x) x + \lambda\theta$

$= \sum_i [y_i - \sigma(\theta^T x_i)] x_i + \lambda\theta$

$= X^T(\sigma(X\theta) - y) + \lambda\theta$

☆ checked soln:

$\nabla^2 \ell = \frac{d}{d\theta} \nabla \ell^T$

$= \sum_i \nabla \theta^{\sigma}(\theta^T x)\, x_i^T + \lambda I$

$= X^T \operatorname{diag}[\sigma(X\theta)(1-\sigma(X\theta))] X + \lambda I$

b) $P(y=c \mid x, W) = \frac{1}{Z}\exp(w_c^T x)$

$= \frac{\exp(w_c^T x)}{\sum_i \exp(w_i^T x)}$

☆ saw soln: Assuming each of W columns have a gaussian distr

$nll(W) = -\log \prod_i \prod_c \mu_{ic}^{y_{ic}} - \lambda\, tr(W^T W)$

$= \sum_i \sum_c y_{ic} \log \mu_{ic} + \lambda\, tr(W^T W)$

where

$\nabla_W nll = X^T(\mu - y) + \lambda W$