B-555 Machine Learning
Written Assignment 3

Ans.2 $f(x) = x_1^3 + 5x_1x_2^2 - 7x_1^2x_2$

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Now, Acc to Newton Raphson method :

$$x_{n+1} = x_n - J(x_0)^{-1}_{|x_0} f(x_0)$$

$$\frac{\partial f}{\partial x} = \begin{pmatrix} 3x_1^2 + 5x_2^2 - 14x_1x_2 \\ 10x_1x_2 - 7x_1^2 \end{pmatrix}$$

$$\text{Hessian} = \begin{pmatrix} 6x_1 - 14x_2 & 10x_2 - 14x_1 \\ 10x_2 - 14x_1 & 10x_1 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 6x_1 - 14x_2 & 10x_2 - 14x_1 \\ 10x_2 - 14x_1 & 10x_1 \end{pmatrix}^{-1} \begin{pmatrix} -6 \\ 3 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \underbrace{\begin{pmatrix} -8 & -4 \\ -4 & 10 \end{pmatrix}}_{A}^{-1} \begin{pmatrix} -6 \\ 3 \end{pmatrix}$$

$A_{11} = 10 \qquad A_{12} = +4$

$A_{21} = +4 \qquad A_{22} = -8$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{-80-16} \begin{pmatrix} 10 & 4 \\ 4 & -8 \end{pmatrix} \begin{pmatrix} -6 \\ 3 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \left(\frac{-1}{96} \begin{pmatrix} 10 & 4 \\ 4 & -8 \end{pmatrix} \begin{pmatrix} -6 \\ 3 \end{pmatrix} \right)$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} -10/96 & -4/96 \\ -4/96 & 8/96 \end{pmatrix} \begin{pmatrix} -6 \\ 3 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 60/96 - 12/96 \\ 24/96 + 24/96 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

$$\boxed{x_1 = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}}$$

Ans. 5. Poisson Likelihood fn., $\text{Poisson}(x|\lambda) = \frac{1}{x!} \times e^{-\lambda} \times \lambda^{x}$

Gamma Prior, $\text{Gamma}(\lambda|a,b) = \frac{1}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$

We know, Gamma Prior to Poisson likelihood is a conjugate prior to the likelihood fn. So, the posterior will also be a gamma distⁿ

Gamma Posterior, $\text{Gamma}(\lambda|a,b,x) \propto$ Prior $\times$ likelihood

To approximate the posterior, we'll use Laplace approximation which approximates λ_{MAP} as the mean of the approximate gaussian distribution. And the $-(\text{Hessian})^{-1}$ as the covariance of the gaussian distⁿ. We'll ignore terms not consisting λ $\because$ they'll differentiate to 0

$f(x) =$ Prior $\times$ Likelihood

$\log f(x) = -\lambda + x \log \lambda + (a-1) \log \lambda - b\lambda$

$$\frac{df}{d\lambda} = -1 + \frac{x}{\lambda} + \frac{a-1}{\lambda} - b = 0$$

$$\Rightarrow \frac{a-1+x}{\lambda} - b - 1 = 0 \quad — (1)$$

For $a=1$ $b=1$ $x=3$

$$\Rightarrow \frac{3}{\lambda} - 2 = 0$$

$$\Rightarrow \boxed{\lambda_{MAP} = \frac{3}{2}}$$

To calculate Hessian diff (1) w.r.t λ again.

$$H = \frac{-(a-1+x)}{\lambda^2}$$

$$\text{Covariance} = -(\text{Hessian})^{-1} = +\left(\frac{\cancel{3}\times 4}{\cancel{9}\;3}\right)^{-1}$$

$$= +\frac{3}{4}$$

$$\therefore P(\lambda \mid x, a, b) = N\left(\frac{3}{2}, +\frac{3}{4}\right)$$

True function:

Posterior $\propto$ Prior $\times$ likelihood

$$= \cancel{\lambda^{a-1} \times \exp\{-b\lambda\} \times \exp\{-\lambda\}}$$

$$= \frac{b^a}{\Gamma a}\lambda^{a-1} \times \exp\{-b\lambda\} \times \frac{1}{x!} \times e^{-\lambda} \times \lambda^{x}$$

$$= \frac{1}{\Gamma(a)\times x!} b^a \times \lambda^{a-1+x} \times \exp\{-\lambda - b\lambda\}$$

Constant Term

$$= \frac{1}{\Gamma(a)\times x!} b^a \times \lambda^{\overbrace{(a+x)}^{A}-1} \times \exp\{-\lambda\overbrace{(b+1)}^{\beta}\}$$

$\therefore$ True fn $=$ Gamma$(\lambda \mid a+x, b+1)$

$=$ Gamma$(\lambda \mid 4, 2)$

The plot for True fn. & lapplace approximate is attached with the PDF. It has the R code attached for the plot

Ans 1. $p(X|\mu,\lambda) = \prod_{i=1}^{N} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left\{\frac{-\lambda}{2}(x_i-\mu)^2\right\}$

Likelihood

Prior: $p(\mu,\lambda) = N\left(\mu|\mu_0, \frac{1}{\beta_0\lambda}\right) \text{Gamma}(\lambda|a_0,b_0)$

$$= \left(\frac{\beta_0\lambda}{2\pi}\right)^{\frac{1}{2}} \times \exp\left\{-\frac{\beta_0\lambda}{2}(\mu-\mu_0)^2\right\} \frac{1}{\Gamma(a_0)} b_0^{a_0} \lambda^{a_0-1} e^{-b_0\lambda}$$

Posterior $\propto$ Prior x Likelihood

$$\propto \underbrace{\left(\frac{\beta_0}{2\pi}\right)^{\frac{1}{2}} \times \frac{1}{\Gamma(a_0)} \times \lambda^{a_0-\frac{1}{2}} \times \exp\left\{\frac{-\beta_0\lambda}{2}(\mu-\mu_0)^2\right\} \times \exp\left\{-\lambda b_0\right\}}_{\text{Prior}}$$

$$\times \underbrace{\left(\frac{\lambda}{2\pi}\right)^{\frac{N}{2}} \times \exp\sum_{i}^{N}\left\{\frac{-\lambda}{2}(x_i-\mu)^2\right\}}_{\text{Likelihood}}$$

$$\text{Post} \propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left\{\frac{-\beta_0\lambda}{2}(\mu-\mu_0)^2 - \lambda b_0 - \frac{\lambda}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right\}$$

$$\propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{\beta_0(\mu-\mu_0)^2 + 2b_0 + \sum_{i=1}^{N}(x_i-\mu)^2\right\}\right]$$

$$\text{Post} \propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{\beta_0(\mu^2+\mu_0^2-2\mu\mu_0)+2b_0+\sum_{i=1}^{N}x_i^2+N\mu^2-2\mu\sum_{i=1}^{N}x_i\right\}\right]$$

$$\propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{\mu^2(\beta_0+N)-2\mu\left(\beta_0\mu_0+\sum_{i=1}^{N}x_i\right)+\mu_0^2\beta_0+2b_0+\sum_{i=1}^{N}x_i^2\right\}\right]$$

$$\propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{(\beta_0+N)\left(\mu^2-2\mu\frac{(\beta_0\mu_0+\sum_{i=1}^{N}x_i)}{\beta_0+N}\right)+\sum_{i=1}^{N}x_i^2+2b_0+\mu_0^2\beta_0\right\}\right]$$

$$\propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{(\beta_0+N)\left(\mu^2-2\mu\frac{(\beta_0\mu_0+\sum_{i=1}^{N}x_i)}{\beta_0+N}+\left(\frac{\beta_0\mu_0+\sum_{i=1}^{N}x_i}{\beta_0+N}\right)^2-\left(\frac{\beta_0\mu_0+\sum_{i=1}^{N}x_i}{\beta_0+N}\right)^2\right)\right.\right.$$
$$\left.\left.+\sum^{N}x_i^2+2b_0+\mu_0^2\beta_0\right\}\right]$$

$$\propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{(\beta_0+N)\left(\mu-\frac{(\beta_0\mu_0+\sum_{i=1}^{N}x_i)}{\beta_0+N}\right)^2+\sum x_i^2+2b_0+\mu_0^2\beta_0\right.\right.$$
$$\left.\left.-\frac{\left(\beta_0\mu_0+\sum_{i=1}^{N}x_i\right)^2}{\beta_0+N}\right\}\right]$$

$$\propto \lambda^{a_0-\frac{1}{2}+\frac{N}{2}} \times \exp\left[\frac{-\lambda}{2}\left\{(\beta_0+N)\left(\mu-\frac{\beta_0\mu_0+\sum_{i=1}^{N}x_i}{\beta_0+N}\right)^2\right\}\right]$$
$$\times \exp\left[-\lambda\left(\frac{\sum x_i^2+\mu_0^2\beta_0}{2}+b_0-\frac{\left(\beta_0\mu_0+\sum_{i=1}^{N}x_i\right)^2}{\beta_0+N}\right)\right]$$

$$\text{Posterior} = \sqrt{\frac{\beta_n}{2\pi}} \times \exp\left\{-\frac{\beta_n\lambda}{2}(\mu-\mu_N)^2\right\} \cdot \frac{1}{\Gamma(a_N)} \cdot b_N^{a_N} \times \lambda^{a_N-\frac{1}{2}} \times \exp\{-b_N\lambda\}$$

Comparing $\boxed{\beta_n = \beta_0+N}$ $\boxed{\mu_N = \frac{\beta_0\mu_0+\sum_{i=1}^{N}x_i}{\beta_0+N}}$

$$b_N = \frac{1}{2}\left(\sum_{i=1}^{N} x_i^2 + \mu_0^2 B_0 - \frac{\left(\sum x_i + B_0\mu_0\right)^2}{B_0 + N}\right)$$

$$\boxed{a_N = a_0 + \frac{N}{2}}$$

Ans. 3.

$$\nabla E(w) = \sum_{n=1}^{N} (y_n - t_n)\phi_n \quad -(4.91)$$

$$(4.108) - E(w_1, \ldots w_k) = -\ln p(T|w_1, \ldots w_K) = -\sum^{N}\sum^{K} t_{nk} \ln y_{nk}$$

$$\nabla_{w_j} E(w_1, \ldots w_K) = \sum^{N} (y_{nj} - t_{nj})\phi_n \quad -(4.109)$$

Now, $E(w_1, \ldots w_K) = -\sum_{n=1}^{N}\sum_{n=1}^{K} t_{nk} \ln(y_{nk})$

$$\frac{dE}{\partial w_j} = -\sum_{n=1}^{N}\sum_{n=1}^{K} t_{nk} \times \frac{1}{y_{nj}} \times y_{nj}(I_{kj} - y_{nj})\phi(x_n)$$

$$= \sum_{n=1}^{N}\sum_{n=1}^{K} t_{nk} y_{nj}\phi(x_n) - \sum_{n=1}^{N}\sum_{n=1}^{K} t_{nk} I_{kj}\phi(x_n)$$

Now, $\sum^{K} t_{nk} = 1$

$$= \sum_{n=1}^{N} y_{nj}\phi(x_n) - \sum_{n=1}^{N} t_{nj}\phi(x_n)$$

$$= \boxed{\sum_{n=1}^{N} (y_{nj} - t_{nj})\phi(x_n)}$$

Ans. 4 (i) $p_2(x) = \frac{1}{s} p_1\left(\frac{x}{s}\right)$

~~Now $p_1(x)$ is normalized probability density fn.~~

$$p_2(x) = \frac{1}{s} \times f\left(\frac{x}{s}\right) \times g(\eta) \times \exp\left\{\eta \frac{x}{s}\right\}$$

$$p_1(x) = f(x) \times g(\eta) \times \exp\{\eta x\}$$

$p_1(x)$ is normalized prob. density fn.

$$p_1(x) = \int f(x) \times g(\eta) \times \exp\{\eta x\} = 1$$

$$p_2(x) = \int \frac{1}{s} f\left(\frac{x}{s}\right) \times g(\eta) \times \exp\left\{\eta \frac{x}{s}\right\}$$

Now, $\frac{x}{s} = u$

$dx = s\,du$

$$= \int \frac{1}{\cancel{s}} f(u)\, g(\eta) \exp\{\eta u\} \cancel{s}\, du$$

$$= \boxed{1}$$

So $p_2(x)$ is also a normalized probability density fn.

(ii) Now, $p(x|s,\eta) = \frac{1}{s} h\left(\frac{x}{s}\right) g(\eta) \times \exp\left\{\frac{\eta x}{s}\right\}$

To prove: θ = Expectatⁿ Parameter $= -s \frac{d}{d\eta} \ln g(\eta)$

$$\cancel{\frac{dp}{d\eta} = \frac{d}{d\eta} \int \frac{1}{s} h\left(\frac{x}{s}\right)}$$

$$\frac{df}{d\eta} = \frac{d}{d\eta} g(\eta) \int \frac{1}{s} h\left(\frac{x}{s}\right) \exp\left\{\frac{\eta x}{s}\right\} + g(\eta) \int \frac{1}{s} h\left(\frac{x}{s}\right) \exp\left\{\frac{\eta x}{s}\right\} \frac{x}{s}$$

$$\frac{df}{d\eta} = \frac{d}{d\eta} g(\eta) \cdot \frac{1}{g(\eta)} + \frac{1}{s} E[x] = 0$$

This is of the form: $\frac{\partial}{\partial\eta} \ln g(\eta) = \frac{1}{g(\eta)} \frac{\partial}{\partial\eta} (g(\eta))$

$\therefore \log g(\eta) + \frac{1}{s} E[x] = 0$

$$\boxed{E[x] = -\frac{\ln g(\eta)}{s}}$$

(iii) To prove: $Var(x|\eta) = -s^2 \frac{d^2}{d\eta^2} \ln g(\eta)$

Now $E[x|\eta, s] = -s \frac{d}{d\eta} \ln g(\eta)$

Differentiating w.r.t η

~~$-s \frac{\partial}{\partial\eta^2} \ln g(\eta) = \frac{\partial}{\partial\eta}\left(\int\right.$~~

$$-s \frac{d^2}{\partial\eta^2} \ln(g(\eta)) = \frac{\partial}{\partial\eta}\left(\int \frac{1}{s} f\left(\frac{x}{s}\right) g(\eta) e^{\eta x/s} x \, dx\right.$$

$$= \frac{d}{d\eta} g(\eta) \int f\left(\frac{x}{s}\right) e^{\eta x/s} \frac{x}{s} dx \, \cancel{+g(\eta)}$$

$$+ g(\eta) \int f\left(\frac{x}{s}\right) e^{\eta x/s} \left(\frac{x^2}{s^2}\right) dx$$