CS 698S

Assignment 2

NAME: PRATIK MISHRA

ROLL: 14493

Ques 1:

We know that, the KL divergence is non-negative.

i.e, $KL(q||p) = -\int q(y) \log\left(\frac{p(y|x)}{q(y)}\right) dy \geq 0$

Now, in the above expression, let.

$q(y) = p(y_*|y)$

and $p(y|x) = p(y_*|\hat{\theta})$

Then,

$$KL(q||p) = -\int p(y_*|y) \log\left(\frac{p(y_*|\hat{\theta})}{p(y_*|y)}\right) dy_*$$

$$= -\int p(y^*|y) \log(p(y_*|\hat{\theta})) dy_*$$

$$+ \int p(y^*|y) \log(p(y_*|y)) dy_*$$

$$= E(\ell(y_*, \mu_1)) - E(\ell(y_*, \mu_2)) \geq 0$$

Hence, $E(\ell(y_*, \mu_2)) \leq E(\ell(y_*, \mu_1))$

Ques 2

$$Gamma(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

In order to approximate this distribution with a normal distribution using Laplace approximation, we need to find the point where Gamma(x|a,b) attains its maximum and set that point as the mean of the Normal distribution.

$$\frac{d}{dx} Gamma(x|a,b) = \frac{b^a}{\Gamma(a)} (a-1) x^{a-2} e^{-bx} - \frac{b^a}{\Gamma(a)} x^{a-1} b e^{-bx}$$

equating it with zero, we get

$$(a-1) x^{a-2} e^{-bx} = x^{a-1} b e^{-bx}$$

$$\Rightarrow \quad x = \frac{a-1}{b}$$

Hence, mean of the Normal distribution $= \frac{a-1}{b}$

To get the same maximum value at $x = \frac{a-1}{b}$

$$\frac{b^a}{\Gamma(a)} \left(\frac{a-1}{b}\right)^{a-1} e^{-(a-1)} = \frac{1}{\sqrt{2\pi}\,\sigma}$$

Standard deviation of Normal distribution

$$\Rightarrow \sigma = \frac{\Gamma(a)}{\sqrt{2\pi}\, b} \left(\frac{e}{a-1}\right)^{a-1}$$

So, $N\left(x \,\middle|\, \frac{a-1}{b}, \frac{\Gamma^2(a)}{2\pi b^2}\left(\frac{e}{a-1}\right)^{2a-2}\right)$ as the

Laplace approximation for Gamma $(x \mid a, b)$

- Mean of Gamma $(x \mid a, b) = a/b$

  Variance of Gamma $(x \mid a, b) = a/b^2$

Thus, Normal approximation with same mean & variance as Gamma distribution would be:

$$N\left(x \,\middle|\, a/b,\ a/b^2\right)$$

For the 2 approximations to be roughly same, we must have,

$$\frac{a-1}{b} \approx \frac{a}{b} \Rightarrow b \to \infty$$

Hence, for large values of b, we can have the 2 approximations as roughly same

Also, equating the variance, we have,

$$\frac{\Gamma^2(a)}{2\pi b^2}\left(\frac{e}{a-1}\right)^{2a-2} \approx \frac{a}{b^2}$$

$$\Rightarrow \Gamma(a) \approx \sqrt{2\pi a}\ (a/e)^{a-1}$$

These 2 conditions would ensure the 2 approximations being roughly equal.

Ques 3

$$P(\mu|x,\beta) = \frac{P(x|\mu,\beta)\cdot P(\mu)}{\int_{-\infty}^{\infty} P(x|\mu\beta)\cdot P(\mu)\, d\mu}$$

$$\propto N(\mu,\beta^{-1})\cdot N(\mu_0, s_0)$$

Because of the property of conjugacy.

$$P(\mu|x,\beta) = N\left(\frac{\left(\frac{\mu_0}{s_0}+\frac{x}{\beta^{-1}}\right)}{\frac{1}{s_0}+\frac{1}{\beta^{-1}}}, \left(\frac{1}{s_0}+\frac{1}{\beta^{-1}}\right)^{-1}\right)$$

$$= N\left(\frac{\mu_0 + s_0\beta x}{1+s_0\beta}, \frac{s_0}{1+s_0\beta}\right)$$

Also,

$$P(\beta|x,\mu) = \frac{P(x|\mu,\beta)\cdot P(\beta)}{\int_{-\infty}^{\infty} P(x|\mu,\beta)\cdot P(\beta)\, d\beta}$$

$$\propto N(\mu,\beta^{-1})\cdot \text{Gamma}(a,b)$$

Because of the property of conjugacy,

$$P(\beta|x,\mu) = \text{Gamma}\left(a+\frac{1}{2}, b+\frac{x-\mu}{2}\right)$$

Ques 4

$$p(\mu, \tau \mid \mu_0, \lambda_0, \alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}\sqrt{\lambda_0}}{\Gamma(\alpha_0)\sqrt{2\pi}} \tau^{\alpha_0 - \frac{1}{2}} \exp(-\beta_0\tau) \exp\left(-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}\right)$$

$$= \frac{\tau^{-1/2}}{\sqrt{2\pi}} \frac{\beta_0^{\alpha_0}\sqrt{\lambda_0}}{\Gamma(\alpha_0)} \exp\left(-\beta_0\tau - \frac{\lambda_0\tau\mu^2}{2} - \frac{\lambda_0\mu_0^2\tau}{2} + \lambda_0\tau\mu_0\mu + \alpha_0\log\tau\right)$$

$$= \frac{\tau^{-1/2}}{\sqrt{2\pi}} \exp\left(-\beta_0\tau - \frac{\lambda_0\tau\mu^2}{2} - \frac{\lambda_0\tau\mu_0^2}{2} + \lambda_0\tau\mu\mu_0 + \alpha_0(\log\tau) - \log(\Gamma(\alpha_0)) + \alpha_0\log(\beta_0) + \frac{1}{2}\log(\lambda_0)\right)$$

Thus,

Sufficient statistics are:

$$\phi(x) = [\tau, \tau\mu^2, \tau\mu, \log\tau]^T$$

Natural parameters are:

$$\theta = \left[-\beta_0 - \frac{\lambda_0\mu_0^2}{2}, \frac{-\lambda_0}{2}, \lambda_0\mu_0, \alpha_0\right]^T$$

Log partition function:

$$A(\theta) = -\log(\Gamma(\alpha_0)) + \alpha_0\log\beta_0 + \frac{1}{2}\log(\lambda_0)$$

Task 5

The model with $k=3$ seems to best ~~[illegible]~~ explain the data because the model with $k=3$ is more confident about the mean as compared to models with $k=1$ and 2. This is visible in the graphs where we can see how close the curves for mean, mean $+2SD$, mean $-2SD$ are in the 3 models.

A new point x^* should ~~[illegible]~~ be added in the region $[-4, -3]$ because in this region, the models are least confident about ~~[illegible]~~ the mean as is visible in the plots of the mean for the 3 models.