HW8

20/06/2024

Q1(a)

$$\begin{aligned}
\mathbb{E}_g(log\ f_{\sigma^2}(X)) &= \mathbb{E}_g\left[log\left(\frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right)\right)\right] \\
&= \mathbb{E}_g\left[-\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X^2 - 2X\mu + \mu^2)\right] \\
&= -\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\mathbb{E}_g[X^2] - 2\mu\mathbb{E}_g[X] + \mu^2\right) \\
&= -\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(Var_g(X) + \mathbb{E}_g[X]^2 - 2\mu m + \mu^2\right) \\
&= -\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\frac{\nu}{\nu-2}s^2 + m^2 - 2\mu m + \mu^2\right)
\end{aligned}$$

(b) As shown in the lecture, minimizing KL divergence $D(g, f_{\sigma^2})$ is equivalent to maximizing $\mathbb{E}_g(log\ f_{\sigma^2}(X))$.

$$\begin{aligned}
\frac{d}{d\sigma^2}\mathbb{E}_g(log\ f_{\sigma^2}(X)) &= -\frac{1}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{1}{2\sigma^4}\left(\frac{\nu}{\nu-2}s^2 + m^2 - 2\mu m + \mu^2\right) \overset{!}{=} 0 \\
\implies \hat{\sigma^2} &= \frac{\nu}{\nu-2}s^2 + m^2 - 2\mu m + \mu^2 \\
&= \frac{\nu}{\nu-2}s^2 \\
\frac{d}{d\sigma^2}\frac{d}{d\sigma^2}\mathbb{E}_g(log\ f_{\sigma^2}(X)) &= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\left(\frac{\nu}{\nu-2}s^2 + m^2 - 2\mu m + \mu^2\right) \\
&= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\left(\frac{\nu}{\nu-2}s^2\right)
\end{aligned}$$

$\implies \mathbb{E}_g(log\ f_{\sigma^2}(X))$ is concave in the range $0 < \sigma^2 < \frac{2\nu}{\nu-2}s^2$, and achieve maximum at $\hat{\sigma^2} = \frac{\nu}{\nu-2}s^2$ given $\nu > 2,\ s > 0$. The result is consistent with the asymptotic normally distributed properties of t-distribution when $\nu \to \infty$, with the same mean and variance.