MA 590 Homework 10

Dane Johnson

June 16, 2020

1

Part a

We want to find $x_{\mathrm{MAP}} = \arg\max\limits_{x \in \mathbb{R}} \pi_{\mathrm{post}}(x) = \arg\min\limits_{x \in \mathbb{R}} -\log\left(\pi_{\mathrm{post}}(x)\right)$. In the case that $\frac{\alpha}{\sigma_0} >> \frac{1-\alpha}{\sigma_1}$, $\pi_{\mathrm{post}}(x) \approx \frac{\alpha}{\sigma_0}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_0^2}\right)$. We will find the minimum of $-\log\left(\pi_{\mathrm{post}}(x)\right)$.

$$0 = \frac{d}{dx}\left[-\log\left(\pi_{\mathrm{post}}(x)\right)\right]$$
$$0 = \frac{d}{dx}\left[-\log\left(\frac{\alpha}{\sigma_0}\frac{1}{\sqrt{2\pi}}\right) + \frac{x^2}{2\sigma_0^2}\right]$$
$$0 = \frac{x}{\sigma_0^2} \implies x = 0 \text{ is a critical point.}$$

Then since in this scenario $\frac{d^2}{dx^2}\left[-\log\left(\pi_{\mathrm{post}}(x)\right)\right] = \frac{1}{\sigma_0^2} > 0$ for any x, $x = 0$ must be a minimum of $-\log\left(\pi_{\mathrm{post}}(x)\right)$. So in this scenario $x_{\mathrm{MAP}} = 0$.

In the case that $\frac{\alpha}{\sigma_0} << \frac{1-\alpha}{\sigma_1}$, $\pi_{\mathrm{post}}(x) \approx \frac{1-\alpha}{\sigma_1}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-1)^2}{2\sigma_1^2}\right)$. We will find the minimum of $-\log\left(\pi_{\mathrm{post}}(x)\right)$.

$$0 = \frac{d}{dx}\left[-\log\left(\pi_{\mathrm{post}}(x)\right)\right]$$
$$0 = \frac{d}{dx}\left[-\log\left(\frac{1-\alpha}{\sigma_1}\frac{1}{\sqrt{2\pi}}\right) + \frac{(x-1)^2}{2\sigma_1^2}\right]$$
$$0 = \frac{x-1}{\sigma_1^2} \implies x = 1 \text{ is a critical point.}$$

Then since in this scenario $\frac{d^2}{dx^2}\left[-\log\left(\pi_{\mathrm{post}}(x)\right)\right] = \frac{1}{\sigma_1^2} > 0$ for any x, $x = 0$ must be a minimum of $-\log\left(\pi_{\mathrm{post}}(x)\right)$. So in this scenario $x_{\mathrm{MAP}} = 1$.

Since $\pi_{\mathrm{post}}(x)$ is the sum of two scaled Gaussian functions (which because of our coefficient assumptions here are strictly positive), one with a mean at $x = 0$ and another with $x = 1$, then if the coefficient of one of these 'pieces' is not significantly larger than the other, the tails of the Gaussian functions interact nontrivially to give an x_{MAP} between 0 and 1. The coefficients then determine whether x_{MAP} is closer to 0 or 1 (see figures 3 and 4).

Next the conditional mean:

$$
\begin{aligned}
x_{\mathrm{CM}} &= \int_{-\infty}^{\infty} \frac{\alpha x}{\sigma_0\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_0^2}\right) + \frac{(1-\alpha)x}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(x-1)^2}{2\sigma_1^2}\right)\,dx \\
&= \alpha\int_{-\infty}^{\infty} \frac{x}{\sigma_0\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_0^2}\right)\,dx + (1-\alpha)\int_{-\infty}^{\infty} \frac{x}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(x-1)^2}{2\sigma_1^2}\right)\,dx \\
&= \alpha(0) + (1-\alpha)(1) = 1 - \alpha \ .
\end{aligned}
$$

The simplification comes from the fact that $\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_0^2}\right)$ and $\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(x-1)^2}{2\sigma_1^2}\right)$ are Gaussian functions and can serve as posterior distributions. In lecture we saw that in this case $x_{\mathrm{MAP}} = \mu$, the mean of the Gaussian. Splitting the integral as we did shows the definition of x_{CM} for these two Gaussian functions, which have means 0 and 1.

Part b

See the end of the document for the figures. In the first case, we might say that x_{MAP} is a better representation of the underlying function since a larger proportion of the area under the curve (probability) is clustered around this value while in the second case the conditional mean is better under the same assumption that we want a value near where a larger proportion of the data is clustered. In the first case we have small variance relative to α, explaining why the data is clustered away from $1 - \alpha$ while in the second case the larger spread around $x = 1$ means that $1 - \alpha \approx 1$ will be nearer to where more area under the curve is located.

In the first case the conditional covariance is

$$\sigma^2 = \alpha \sigma_0^2 + (1 - \alpha)(\sigma_1^2 + 1) - (1 - \alpha)^2 = 0.254$$

In the second case the conditional covariance, using the same formula, is $\sigma^2 = 0.0198$. In both cases the conditional covariance is much closer to α than either of σ_0^2 or σ_1^2.

2

The model can be transformed to

$$\log(Y) = \log(E) + \log(G(X)) \,,$$

where $\log(E) \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. When the variance is unknown the likelihood function for $\log(Y)$ conditioned on $X = x$ is

$$\pi(\log(y) \mid x, \sigma^2) = \frac{1}{(2\pi)^{m/2}\sigma^m} \exp\left(-\frac{1}{2\sigma^2}||\log(y) - \log(G(X))||_2^2\right) \,,$$

while in the case that the variance is known the likelihood function is then

$$\pi(\log(y) \mid x, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}||\log(y) - \log(G(X))||_2^2\right) \,.$$

Figures