CMPUT 466 Assignment 2 Oct. 26

Jinzhu Li 1461696

Question 1:

(a) The MAP estimate of the parameter θ is:

$$\theta_{map} = \arg\max_{\theta} \{ P(D|\theta) P(\theta) \}$$

θ itself follows a normal distribution $N(\mu, \sigma^2)$,

$$P(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\theta-\mu)^2}$$

$$\log P(\theta) = -\log\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(\theta-\mu)^2$$

$$\frac{d}{d\theta}\log P(\theta) = -\frac{1}{\sigma^2}(\theta-\mu)$$

$$P(D|\theta) = \prod_{i=1}^{n} P(x_i|\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x_i-\theta)^2}$$

$$\log P(D|\theta) = \sum_{i=1}^{n}\left[-\log\sqrt{2\pi\sigma_0^2} - \frac{1}{2\sigma_0^2}(x_i-\theta)^2\right]$$

$$= -\sum_{i=1}^{n}\log\sqrt{2\pi\sigma_0^2} - \sum_{i=1}^{n}\frac{1}{2\sigma_0^2}(x_i-\theta)^2$$

Take the logarithm of θmap.

$$\log \theta_{map} = \log P(D|\theta) + \log P(\theta)$$

$$= -\sum_{i=1}^{n}\log\sqrt{2\pi\sigma_0^2} - \sum_{i=1}^{n}\frac{1}{2\sigma_0^2}(x_i-\theta)^2 + \log P(\theta)$$

$$\frac{d}{d\theta}\log\theta_{map} = \sum_{i=1}^{n}\frac{1}{\sigma_0^2}(x_i-\theta) + \frac{d}{d\theta}\log P(\theta)$$

$$= \sum_{i=1}^{n}\frac{1}{\sigma_0^2}(x_i-\theta) - \frac{1}{\sigma^2}(\theta-\mu)$$

let $\frac{d}{d\theta}\log\theta_{map} = 0 \Rightarrow \frac{\sum_{i=1}^{n}(x_i-\theta)}{\sigma_0^2} = \frac{\theta-\mu}{\sigma^2}$

$$\frac{\sum_{i=1}^{n} x_i - n\theta}{\sigma_0^2} = \frac{\theta-\mu}{\sigma^2}$$

$$\sigma^2\sum_{i=1}^{n} x_i - \sigma^2 n\theta = \sigma_0^2\theta - \sigma_0^2\mu$$

$$\theta(-\sigma_0^2 - \sigma^2 n) = -\sigma_0^2\mu - \sigma^2\sum_{i=1}^{n} x_i$$

$$\Rightarrow \boxed{\theta = \frac{\sigma_0^2 \mu + \sigma^2 \sum_{i=1}^{n} x_i}{\sigma_0^2 + \sigma^2 n}} \quad \circledast$$

Further calculation for $\frac{d^2}{d^2\theta} \log \theta_{map} = -\frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} < 0$

Therefore, $\circledast$ is a local maximum point.

ie the MAP estimate of θ.

(b) Assume $\mu = 0$, we have, $p(\theta) = \frac{1}{2b} e^{\frac{-|\theta|}{b}}$

For $\theta > 0$, $p(\theta) = \frac{1}{2b} e^{-\frac{\theta}{b}}$

$$\theta_{MAP} = \arg\max_{\theta} \{ P(D|\theta) P(\theta) \}$$

$$\log \theta_{MAP} = \log P(D|\theta) + \log P(\theta)$$

$$= -\sum_{i=1}^{n} \log \sqrt{2\pi\sigma_0^2} - \sum_{i=1}^{n} \frac{1}{2\sigma_0^2} (x_i - \theta)^2 - \log 2b - \frac{\theta}{b}$$

$$\frac{d}{d\theta} \theta_{MAP} = \sum_{i=1}^{n} \frac{1}{\sigma_0^2} (x_i - \theta) - \frac{1}{b} = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{n} x_i - n\theta}{\sigma_0^2} = \frac{1}{b}$$

$$\Rightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sigma_0^2}{nb}$$

For $\theta < 0$, $p(\theta) = \frac{1}{2b} e^{\frac{\theta}{b}}$

$$\log \theta_{MAP} = \log P(D|\theta) + \log P(\theta)$$

$$= -\sum_{i=1}^{n} \log \sqrt{2\pi\sigma_0^2} - \sum_{i=1}^{n} \frac{1}{2\sigma_0^2} (x_i - \theta)^2 - \frac{1}{2b} + \frac{\theta}{b}$$

$$\frac{d}{d\theta} \theta_{MAP} = \sum_{i=1}^{n} \frac{1}{\sigma_0^2} (x_i - \theta) + \frac{1}{b} = 0$$

$$\Rightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sigma_0^2}{nb}$$

Therefore, we do not have a closed form solution for θ_{MAP}

This objective is continous but not smooth, i.e. is not differentiable when $\theta = 0$, making the expression of $\theta = \frac{\sum_{i=1}^{n} x_i}{n} + \left| \frac{\sigma_0^2}{nb} \right|$ non-differentiable.

We can use alternative methods; for smooth component ($\theta \neq 0$) use gradient descent, and for values in θ that are close to zero, use subgradient descent.

In this case, the derivative of θ should between the derivative of left and right side of θ, in this question, the absolute value term will be 0

$$\mathrm{Prox}_{\eta\lambda}(\theta) = \begin{cases} \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sigma_0^2}{nb} - \eta\lambda & \text{if } \theta > \eta\lambda \\ \frac{\sum_{i=1}^{n} x_i}{n} & \text{if } |\theta| < \eta\lambda \\ \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sigma_0^2}{nb} + \eta\lambda & \text{if } \theta < \eta\lambda \end{cases}$$

for some η and λ.

(c) $\theta_{MAP} = \underset{\theta}{\arg\max} \{ P(D|\theta) P(\theta) \}$

$$P(D|\theta) = \prod_{i=1}^{n} P(x_i | \theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \Sigma_0^2}} e^{-\frac{1}{2\Sigma_0^2}(x_i-\theta)^2}$$

$$\log P(D|\theta) = \sum_{i=1}^{n} -\log\sqrt{2\pi\Sigma_0^2} - \sum_{i=1}^{n} \frac{1}{2\Sigma_0^2}(x_i-\theta)^2$$

$$P(\theta) = \prod_{j=1}^{d} P(\theta_j)$$

$$= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma^4}} e^{-\frac{1}{2\sigma^4}\theta_j^2}$$

$$\log P(\theta) = \sum_{j=1}^{d} \log -\sqrt{2\pi\sigma^4} - \sum_{j=1}^{d} \frac{1}{2\sigma^4}\theta_j^2$$

$$\log \theta_{MAP} = \log P(D|\theta) + \log P(\theta)$$

$$= \sum_{i=1}^{n} -\log\sqrt{2\pi\Sigma_0^2} - \sum_{i=1}^{n} \frac{1}{2\Sigma_0^2}(x_i-\theta)^2 - \sum_{j=1}^{d} \log\sqrt{2\pi\sigma^4} - \sum_{j=1}^{d}\frac{1}{2\sigma^4}\theta_j^2$$

since $\theta \in \mathbb{R}^d$

$$\frac{d}{d\theta_j} \log\theta_{MAP} = \sum_{i=1}^{n} \frac{1}{\Sigma_0^2}(x_i - \theta_j) - \frac{1}{\sigma^4}\theta_j = 0$$

Also, $\Sigma_0 = I \in \mathbb{R}^{d\times d}$, thus $\Sigma_0^2 = 1$

$$\sum_{i=1}^{n}(x_i - \theta_j) - \frac{1}{\sigma^4}\theta_j = 0$$

$$\sum_{i=1}^{n} x_i - n\theta_j - \frac{1}{\sigma^4}\theta_j = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i = \left(n + \frac{1}{\sigma^4}\right)\theta_j = \frac{n\sigma^4+1}{\sigma^4}\theta_j$$

$$\Rightarrow \theta_j = \frac{\sigma^4 \sum_{i=1}^{n} x_i}{n\sigma^4+1} = \theta_{MAP}$$

The MAP for estimate of θ is $\boxed{\theta_{MAP} = \frac{\sigma^4 \sum_{i=1}^{n} x_i}{n\sigma^4+1}}$