Q-1

$X \sim \text{Gaussian}(\mu, \sigma^2)$; μ unknown, σ^2 known

M.L estimate →

$$P(\text{data}|\mu) = JL = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{(2\pi)^{n/2}\cdot\sigma^n} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\log(JL) = -\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} + \text{constant}$$

$$\frac{d}{d\mu}\log(JL) = \sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2} = 0$$

$$\Rightarrow \boxed{\hat{\mu}^{ML} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}}$$

MAP estimate with gaussian prior →

Prior : $P(\mu) = \text{Gaussian}(\mu_0, \sigma_0^2)$; $\mu_0 = 10.5$, $\sigma_0 = 1$

$$\underset{\mu}{\arg\max}\ P(\mu|\text{data}) = \underset{\mu}{\arg\max}\ P(\text{data}|\mu)\cdot P(\mu)$$

$$= \underset{\mu}{\arg\max}\left(\prod_{i=1}^{n} G(x_i;\mu,\sigma^2)\right)\left(G(\mu;\mu_0,\sigma_0^2)\right)$$

$$= \underset{\mu}{\arg\max}\left(\sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)$$

$$= \arg\min_{\mu} \underbrace{\left(\sum_{i=1}^{n} \frac{(\mu - x_i)^2}{2\sigma^2} + \frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right)}_{J(\mu) \text{ (let)}}$$

So, $$\frac{d}{d\mu} J(\mu) = 0 = \sum_{i=1}^{n} \frac{\mu - x_i}{\sigma^2} + \frac{\mu - \mu_0}{\sigma_0^2}$$

$$= \frac{n\mu - n\bar{x}}{\sigma^2} + \frac{\mu - \mu_0}{\sigma^2} \quad ; \quad \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$\Rightarrow \boxed{\hat{\mu}^{MAP1} = \frac{\mu_0 \sigma^2/n + \bar{x}\sigma_0^2}{\sigma^2/n + \sigma_0^2}}$$

MAP estimate with uniform prior :

Prior : $$P(\mu) = \frac{1}{11.5 - 9.5} = \frac{1}{2} \quad \text{if } \mu \in [9.5, 11.5]$$
$$= 0 \quad \text{otherwise.}$$

$$\arg\max_{\mu} P(\mu | \text{data}) = \arg\max_{\mu} P(\text{data}|\mu) \cdot P(\mu)$$

$$= \arg\max_{\mu} \left(\prod_{i=1}^{n} G(x_i ; \mu, \sigma^2) \right) \cdot \frac{1}{2}$$

So, Posterior is maximized when the likelihood is maximized ,

So, $\underline{\hat{\mu}^{MAP2} = \hat{\mu}^{ML} = \bar{x}}$ $\quad$ $\underline{\text{if } \bar{x} \in [9.5, 11.5]}$

But if $\bar{x} < 9.5$, the posterior is a decreasing function between $\mu = 9.5$ and 11.5, so its maximum value occurs at 9.5.

Similarly, if $\bar{x} > 11.5$, the posterior is an increasing function between $\mu = 9.5$ and 11.5, so it's maximum value occurs at 11.5.

Hence,
$$\hat{\mu}^{MAP2} = \begin{cases} 9.5 & \text{if } \bar{x} < 9.5 \\ \bar{x} & \text{if } \bar{x} \in [9.5, 11.5] \\ 11.5 & \text{if } \bar{x} > 11.5 \end{cases}$$

where $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$

Interpretations from graph –

i) As n increases, error decreases and approaches 0 for large values of n.

ii) We would prefer MAP 1 estimate (with gaussian prior) because it gives lower absolute errors than other estimates, even when n is lower.