Date:

1. Let us consider a dataset X of N values which are drawn from a univariate Gaussian with given variance σ^2_{true} and unknown mean μ

$$P(X|\mu,\sigma_{true}) = \frac{1}{(\sigma_{true}\sqrt{2\pi})^n} \exp\left(\frac{-1}{2\sigma^2_{true}} \sum_i (x_i-\mu)^2\right)$$

Clearly, $P(X|\mu,\sigma_{true}) \equiv P(X|\mu)$

$$\Rightarrow \quad \log P(X|\mu) = -n\log(\sigma_{true}\sqrt{2\pi}) - \frac{1}{2\sigma^2_{true}} \sum_i (x_i-\mu)^2$$

To find $\hat{\mu}_{MLE}$, we minimize $\Sigma(x_i-\mu)^2$.

Hence, we find that

$$\frac{d}{d\mu} \Sigma (x_i-\mu)^2 \Big|_{\hat{\mu}_{MLE}} = 0$$

Thus,

$$\boxed{\hat{\mu}_{MLE} = \frac{\Sigma x_i}{n}}$$

- Now, we have the Gaussian Prior on μ:

$$\Rightarrow \quad P(\mu) = \frac{1}{\sigma_{prior}\sqrt{2\pi}} \exp\left(\frac{-(\mu_0-\mu)^2}{2\sigma^2_{prior}}\right)$$

Thus, $$P(\mu|X) = \frac{P(X|\mu)P(\mu)}{\underbrace{\int_{-\infty}^{\infty} P(X|\mu)P(\mu)\,d\mu}_{\text{Normalizing constant}}}$$

Hence,
$$P(\mu|X) \propto \frac{1}{(\sigma_{true}\sqrt{2\pi})^{N}(\sigma_{prior}\sqrt{2\pi})} \exp\left(\frac{-\Sigma(x_i-\mu)^2}{2\sigma_{true}^2} - \frac{(\mu_0-\mu)^2}{2\sigma_{prior}^2}\right)$$

$\hat{\mu}_{MAP1}$ minimizes the above value.

~~We~~ For the minimization, we need to ~~maximize~~ differentiate $\frac{\Sigma(x_i-\mu)^2}{2\sigma_{true}^2} + \frac{(\mu_0-\mu)^2}{2\sigma_{prior}^2}$ and solve with 0.

Hence,
$$\frac{1}{\sigma_{true}^2}\Sigma(\hat{\mu}_{MAP1}-x_i) + \frac{(\hat{\mu}_{MAP1}-\mu_0)}{\sigma_{prior}^2} = 0.$$

$$\Rightarrow \left(\frac{n}{\sigma_{true}^2} + \frac{1}{\sigma_{prior}^2}\right)\hat{\mu}_{MAP1} = \frac{\Sigma x_i}{\sigma_{true}^2} + \frac{\mu_0}{\sigma_{prior}^2}$$

Hence,
$$\boxed{\hat{\mu}_{MAP1} = \frac{\Sigma x_i/\sigma_{true}^2 + \mu_0/\sigma_{prior}^2}{n/\sigma_{true}^2 + 1/\sigma_{prior}^2}}$$

$\therefore$ if $\hat{\mu}_{ML} < a$:

$\hat{\mu}_{MAP2} = a$

else if $\hat{\mu}_{MAP2} > b$:

$\hat{\mu}_{MAP2} = b$

else

$$\boxed{\hat{\mu}_{MAP2} = \hat{\mu}_{ML}}$$

Page No.:
Date:

Interpretation:

1. As the sample size N increases, MAP1 and MAP2 estimates tend to the MLE i.e more accuracy

2. MAP1 is made up with an informative prior, but a biased one. Thus for small ~~values~~, samples, the error is small and doesn't vary much.

3. MAP2 just puts a bound on the error of the MLE, only for small sample sizes.

4. As N increases, MLE gets more accurate.

Hence, MAP1 for small samples
MAP2/MLE for large samples.