1.

$$\hat{\sigma}_m^2 = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\hat{\mu}_m)^2$$

$$\text{Bias}(\hat{\sigma}_m^2) = E(\hat{\sigma}_m^2) - \sigma_m^2$$

$$= E\left[\frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\hat{\mu}_m)^2\right] - \sigma_m^2 \quad -①$$

Using $E\left[\sum_{i=1}^{m}(x^{(i)}-\mu_m)^2\right] = (m-1)\cdot\sigma^2$ -

we have from ①

$$\frac{1}{m}E\left[\sum_{i=1}^{m}(x^{(i)}-\mu_m)^2\right] - \hat{\sigma}_m^2$$

$$= \frac{1}{m}\times(m-1)\times\sigma_m^2 - \hat{\sigma}_m^2$$

$$= \frac{-\hat{\sigma}_m^2}{m}$$

Since the difference between expectation and true value is not zero, the estimator $\hat{\sigma}_m^2$ is biased.

$$\tilde{\sigma}_m^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x^{(i)}-\hat{\mu}_m)^2$$

Using the bias definition and hint as above -

$$\text{Bias}(\tilde{\sigma}_m^2) = \frac{m-1\times\tilde{\sigma}_m^2}{m-1} - \sigma_m^2$$

$$= 0$$

$\tilde{\sigma}_m^2$ is unbiased as the difference between expectation & true value is zero.

2. $MSE(\hat{\theta}_m) = E[(\hat{\theta}_m - \theta_m)^2]$

$$= E[(\hat{\theta}_m^2 + \theta_m^2 - 2\hat{\theta}_m\theta_m)] = E[(\hat{\theta}_m)^2] + \theta_m^2 - 2E[\hat{\theta}_m]\theta_m \quad — ①$$

Adding and subtracting $E^2[\hat{\theta}_m]$, we have from ①

$$E[(\hat{\theta}_m)^2] + \theta_m^2 - 2E[\hat{\theta}_m]\theta_m + E^2[\hat{\theta}_m] - E^2[\hat{\theta}_m]$$

$$= \underbrace{E^2[\hat{\theta}_m] + \theta_m^2 - 2E[\hat{\theta}_m]\theta_m}_{\downarrow} + \underbrace{E[(\hat{\theta}_m)^2] - E^2[\hat{\theta}_m]}_{\downarrow}$$

$$Bias(\hat{\theta}_m)^2 \quad + \quad Var(\hat{\theta}_m)$$

Hence,

$$MSE(\hat{\theta}_m) = Bias(\hat{\theta}_m)^2 + Var(\hat{\theta}_m).$$

3.

3. $P(Y|X,\theta) \rightarrow$ Applying the Bayes Rule and neglecting the normalizing term $\rightarrow$

$\rightarrow P(Y|X,\theta)\, P(\theta|X)$
↳ not dependent

Hence, $P(Y|X,\theta)\, P(\theta)$

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} \left[\sum_{i=1}^{n} \log \{ p(y^{(i)}|x^{(i)}, \theta) + \log p(\theta) \right]$$