Student Name: Erin Sarlak

Collaboration Statement:

Total hours spent: 15

I discussed ideas with these individuals:

- Patrick Feeney

I consulted the following resources:

- Linear Algebra Cookbook

By submitting this assignment, I affirm this is my own original work that abides by the course collaboration policy.

Links: [HW2 instructions] [collab. policy]

Contents

1a: Problem Statement

Compute the expected value of estimator $\hat{\sigma}^2(x_1, \dots x_N)$, where

$$\hat{\sigma}^2(x_1, \dots x_N) = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{\text{true}})^2 \tag{1}$$

1a: Solution

We start with:

$$E[\sigma^2] = E\left[\frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{\text{true}})^2\right] \tag{2}$$

By the linearity of expectation, we have:

$$= \frac{1}{N} \sum_{n=1}^{N} E[(x_n - \mu_{\text{true}})^2] \tag{3}$$

Expanding the square inside the expectation, we obtain:

$$= \frac{1}{N} \sum_{n=1}^{N} E[(x_n^2 - 2x_n\mu_{\text{true}} + \mu_{\text{true}}^2)] \tag{4}$$

Recognizing that μ_{true} is constant with respect to the expectation, we can factor it out:

$$= \frac{1}{N} \sum_{n=1}^{N} E[x_n^2] - 2\mu_{\text{true}} E[x_n] + \mu_{\text{true}}^2 \tag{5}$$

Using the linearity of expectation, we split it into two terms:

$$= \frac{1}{N} \sum_{n=1}^{N} (E[x_n^2] - 2\mu_{\text{true}} E[x_n]) + \mu_{\text{true}}^2 \tag{6}$$

Then, we substitute $E[x_n^2] = \mu_{\text{true}}^2 + \sigma_{\text{true}}^2$ and $E[x_n] = \mu_{\text{true}}$:

$$= \frac{1}{N} \sum_{n=1}^{N} (\mu_{\text{true}}^2 + \sigma_{\text{true}}^2 - 2\mu_{\text{true}}^2) + \mu_{\text{true}}^2 \tag{7}$$

Simplifying, we get:

$$= \sigma^2_{\text{true}} \tag{8}$$

1b: Problem Statement

Using your result in 1a, explain if the estimator $\hat{\sigma}^2$ is biased or unbiased. Explain why this differs from the biased-ness of the maximum likelihood estimator for the variance, using a justification that involves the mathematical definition of each estimator. (Hint: Why would one be lower than the other?).

1b: Solution

The estimator $\hat{\sigma}^2$ is unbiased because its expected value equals to σ_{true}. On the other hand, the maximum likelihood estimator for the variance is biased because its expected value takes the form of $E[\sigma^2(x_1...x_N)] = \frac{N-1}{N} \cdot \sigma^2_{true} \neq \sigma_{true}$. Therefore, the estimated variance does not match the truth especially for small sample sizes.

2a: Problem Statement

Suppose you are told that a vector random variable $x \in \mathbb{R}^M$ has the following log PDF function:

$$\log p(x) = \text{c} - \frac{1}{2}x^T Ax + b^T x \tag{9}$$

where A is a symmetric positive definite matrix, b is any vector, and c is any scalar constant.

Show that x has a multivariate Gaussian distribution.

2a: Solution

$$= constant + \frac{1}{2}(x^T - \mu^T)S(x - \mu) \tag{10}$$

Distribute the multiplication over the subtraction:

$$= constant + \frac{1}{2}(x^T - \mu^T)S(x - \mu) \tag{11}$$

Expand using the distributive property of matrix multiplication:

$$= constant + \frac{1}{2}(x^T Sx - x^T S\mu - \mu^T Sx + \mu^T S\mu) \tag{12}$$

Since S is symmetric, we know that $x^T S\mu = \mu^T Sx$, and we can group terms differently:

$$= constant + \frac{1}{2}(x^T Sx - 2(x^T S\mu) + \mu^T S\mu) \tag{13}$$

Let: $c = constant + \mu^T S\mu$, combining the constant terms.

$$= c + \frac{1}{2}(x^T Sx - 2(x^T S\mu)) \tag{14}$$

Let:

1. $b = S\mu$
2. $A = S$

, then we get:

$$= c + \frac{1}{2}x^T Sx - b^T x \tag{15}$$

We found the values for A, b, and C that recovers the formula for the multivariate Gaussian, and , by showing they are equal, we proved that x has a multivariate Gaussian.

3a: Problem Statement

Show that we can write $S_{N+1}^{-1} = S_N^{-1} + vv^T$ for some vector $v \in \mathbb{R}^M$.

3a: Solution

$$S_{N+1}^{-1} = S_N^{-1} + vv^T \tag{16}$$

Given that $S_N^{-1} = \alpha I_M + \beta \Phi_{1:N}^T \Phi_{1:N}$, when we add one extra feature-response pair (ϕ_{N+1}, t_{N+1}), the S_{N+1}^{-1} becomes:

$$S_{N+1}^{-1} = \alpha I_M + \beta \Phi_{1:N+1}^T \Phi_{1:N+1} \tag{17}$$

Subtracting S_N^{-1} from S_{N+1}^{-1} and then simplifying the expression results in:

$$S_{N+1}^{-1} - S_N^{-1} = (\alpha I_M + \beta \Phi_{1:N+1}^T \Phi_{1:N+1}) - (\alpha I_M + \beta \Phi_{1:N}^T \Phi_{1:N}) \tag{18}$$

$$\tag{19}$$

$$= \beta \cdot (\Phi_{1:N+1}^T \Phi_{1:N+1} - \Phi_{1:N}^T \Phi_{1:N}) \tag{20}$$

$$\tag{21}$$

Here, $\Phi_{1:N+1}^T \Phi_{1:N+1}$ an be expanded as:

$$\Phi_{1:N+1}^T \Phi_{1:N+1} = \Phi_{1:N}^T \Phi_{1:N} + \boldsymbol{\phi}_{N+1}^T \boldsymbol{\phi}_{N+1} \tag{22}$$

Substitute this in (20):

$$= \beta \cdot (\Phi_{1:N}^T \Phi_{1:N} + \boldsymbol{\phi}_{N+1} \boldsymbol{\phi}_{N+1}^T - \Phi_{1:N}^T \Phi_{1:N}) \tag{23}$$

$$\tag{24}$$

$$= \beta \boldsymbol{\phi}_{N+1} \boldsymbol{\phi}_{N+1}^T \tag{25}$$

Expression (25) could be turned into vv^T:

$$v = \sqrt{\beta} \phi_{N+1} \tag{26}$$

3b: Problem Statement

Next, consider the following identity, which holds for any invertible matrix A:

$$(A + vv^T)^{-1} = A^{-1} - \frac{(A^{-1}v)(v^T A^{-1})}{1 + v^T A^{-1} v} \tag{27}$$

Substitute $A = S_N^{-1}$ and v as defined in 3a into the above. Simplify to write an expression for S_{N+1} in terms of S_N.

3b: Solution

Substitute $A = S_N^{-1}$ and $v = \sqrt{\beta} \boldsymbol{\phi}_{N+1}$ in (27): (note that: ß$(S_N^{-1})^{-1} = S_N)$

$$(S_N^{-1} + vv^T)^{-1} = S_N - \frac{(S_N \sqrt{\beta} \phi_{N+1})((\sqrt{\beta} \phi_{N+1})^T S_N)}{1 + (\sqrt{\beta} \phi_{N+1})^T S_N (\sqrt{\beta} \phi_{N+1})} \tag{28}$$

Since $S_{N+1}^{-1} = S_N^{-1} + vv^T$, the $LHS = S_{N+1}$:

$$S_{N+1} = S_N - \frac{(S_N \sqrt{\beta}\phi_{N+1})((\sqrt{\beta}\phi_{N+1})^T S_N)}{1 + (\sqrt{\beta}\phi_{N+1})^T S_N (\sqrt{\beta}\phi_{N+1})} \tag{29}$$

3c: Problem Statement

Finally, combine your result from 3b with the following two facts

- (i) $\sigma_{N+1}^2(x_*) - \sigma_N^2(x_*) = \phi(x_*)^T \left[S_{N+1} - S_N\right] \phi(x_*)$
- (ii) S_N must be positive definite

to show that

$$\sigma_{N+1}^2(x_*) \leq \sigma_N^2(x_*) \tag{30}$$

3c: Solution

$$\sigma_{N+1}^2(x_*) - \sigma_N^2(x_*) = \phi(x_*)^T \left[S_{N+1} - S_N\right] \phi(x_*) \tag{31}$$

Expand the expression:

$$= \phi(x_*)^T \phi(x_*) \cdot S_{N+1} - \phi(x_*)^T \phi(x_*) \cdot S_N \tag{32}$$

Substitute (3b) in (32):

$$\phi(x_*)^T \phi(x_*) \cdot S_N - \phi(x_*)^T \phi(x_*) \cdot \frac{(S_N \sqrt{\beta}\phi_{N+1})((\sqrt{\beta}\phi_{N+1})^T S_N)}{1 + (\sqrt{\beta}\phi_{N+1})^T S_N (\sqrt{\beta}\phi_{N+1})} - \phi(x_*)^T \phi(x_*) \cdot S_N \tag{33}$$

$$-\phi(x_*)^T \phi(x_*) \cdot \frac{(S_N \sqrt{\beta}\phi_{N+1})((\sqrt{\beta}\phi_{N+1})^T S_N)}{1 + (\sqrt{\beta}\phi_{N+1})^T S_N (\sqrt{\beta}\phi_{N+1})} \tag{34}$$

Because the following are true, this epression is strictly less than or equal to zero:

1. β is positive.
2. S_N is positive definite.
3. $\phi_{N+1}^T \phi_{N+1}$ has to be positive.

4. Because of bullet points (2) and (3), the $(\sqrt{\beta}\phi_{N+1})^T S_N (\sqrt{\beta}\phi_{N+1})$ is positive as well.

Thus, expression (34) is ≤ 0.