Student Name: Pengcheng Xu

Collaboration Statement:

Total hours spent: 6 hours

I discussed ideas with these individuals:

- I did it on my own
- ...

I consulted the following resources:

- I did it on my own
- ...

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 $\hat{\sigma}^2(x_1, \ldots x_N)$, where

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

1a: Solution

Next, we'll conclude our result step-by-step, and also comment the basis of our derivation on each line:

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

/* Expanding items within parentheses */

$$= E(\frac{1}{N}\sum_{n=1}^{N}(x_n^2 - 2\mu_{true}x_n + u_{true}^2))$$

$$= \frac{1}{N}\sum_{n=1}^{N}(E(x_n^2) - 2E(\mu_{true}x_n) + E(u_{true}^2))$$

/* μ_{true} is a constant, x_n is i.i.d from $N(\mu_{true}, \sigma_{true}^2)$, and $E(x_n^2) = \mu_{true}^2 + \sigma_{true}^2$ */

$$= \frac{1}{N}\sum_{n=1}^{N}(E(x_n^2) - 2\mu_{true}E(x_n) + u_{true}^2)$$

$$= \frac{1}{N}\sum_{n=1}^{N}(E(x_n^2) - u_{true}^2)$$

$$= \frac{1}{N}(\sum_{n=1}^{N}E(x_n^2) - N \cdot u_{true}^2) \tag{eq1}$$

$$= \frac{1}{N}(\sum_{n=1}^{N}(\mu_{true}^2 + \sigma_{true}^2) - N \cdot u_{true}^2) = \sigma_{true}^2$$

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 we computed in 1a is unbiased (since its expectation is σ^2_{true}). The reason of the difference between $\hat{\sigma}^2$ and $\hat{\sigma}^2_{ML}$ is as the following:

For $\hat{\sigma}^2$ (i.e. 1a), we could simplify it to (i.e. see (eq1) in 1a):

$$
\begin{aligned}
E(\hat{\sigma}^2) &= \frac{1}{N}(\sum_{n=1}^{N} E(x_n^2) - N \cdot \mu^2_{true}) \\
&= \frac{1}{N}(\sum_{n=1}^{N} E(x_n^2)) - \cdot \mu^2_{true} \qquad (eq2)
\end{aligned}
$$

For $\hat{\sigma}^2_{ML}$ (i.e. ML estimator), we could simplify it to (i.e. see the slide from day5, page 9):

$$
E(\hat{\sigma}^2_{ML}) = \frac{1}{N}(\sum_{n=1}^{N} E(x_n^2)) - \cdot E(\mu^2_{ML}) \qquad (eq3)
$$

The difference is the last term, where we get $E(\mu^2_{ML})$ in ML estimator, which we could further infer as follows (using $\hat{\mu}_{ML} = \frac{1}{N} \cdot (\sum_{n=1}^{n=N} x_n)$):

$$
\begin{aligned}
E(\mu^2_{ML}) &= \frac{1}{N^2} \cdot E(\sum_{i=1}^{i=N} x_i \cdot \sum_{j=1}^{j=N} x_j) \\
&= \frac{1}{N^2} \cdot E(\sum_{n=1}^{N} x_n^2 + \sum_{i \neq j} x_i x_j)
\end{aligned}
$$

$$= \frac{1}{N^2} \cdot (\sum_{n=1}^{N} E(x_n^2) + \sum_{i \neq j} E(x_i x_j))$$

/* The 2nd expression (i.e. $x_i x_j$, $i \neq j$) has $N^2 - N$ items, and x_n is i.i.d*/

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

$$= \mu_{true}^2 + \frac{1}{N} \cdot \sigma_{true}^2$$

Now, we could plug this into (eq3), then get:

$$E(\hat{\sigma}_{ML}^2) = \frac{1}{N}(\sum_{n=1}^{N} E(x_n^2)) - \cdot \mu_{true}^2 - \frac{1}{N} \cdot \sigma_{true}^2 \tag{eq4}$$

Finally, if we compare (eq2) and (eq4), we can see why $\hat{\sigma}_{ML}^2$ estimator is lower than $\hat{\sigma}^2$ estimator (since ML estimator minus an extra $\frac{1}{N} \cdot \sigma_{true}^2$ item).

2a: Problem Statement

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

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

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

We know that the log of the probability of multi Gaussian distribution has the form:

$$const - \frac{1}{2}(x - \mu)^T S (x - \mu) \tag{2a}$$

where const is a constant with respect to x.

First, we'll claim that if we define $S := A$, $\mu := A^{-1}b$, then equation (2a) could be converted to equation (2), which will show that x has a multivariate Gaussian distribution.

Next, we'll show this conversion step-by-step.

$$const - \frac{1}{2}(x-\mu)^T S(x-\mu) = const - \frac{1}{2}(x - A^{-1}b)^T A(x - A^{-1}b)$$

/* Expanding order-2 multiplication, A is symmetric and pos-definite, so $A^T = A$ and $(A^{-1})^T = (A^T)^{-1}$ */

$$= const - \frac{1}{2}x^T Ax + \frac{1}{2}b^T(A^{-1})^T Ax + \frac{1}{2}x^T AA^{-1}b - \frac{1}{2}b^T(A^{-1})^T AA^{-1}b$$

$$= const - \frac{1}{2}x^T Ax + \frac{1}{2}b^T(A^T)^{-1}Ax + \frac{1}{2}x^T b - \frac{1}{2}b^T(A^T)^{-1}b$$

$$= const - \frac{1}{2}x^T Ax + b^T x + -\frac{1}{2}b^T A^{-1}b$$

/* The last item $-\frac{1}{2}b^T A^{-1}b$ is constant with respect to x, so it could merge with const to form a new constant */

$$= const^{'} - \frac{1}{2}x^T Ax + b^T x \tag{2b}$$

Where $const^{'} := const - \frac{1}{2}b^T A^{-1}b$.

Now, we show that log probability of x (i.e. (2)) could be written in the form of equation (2a), so x has a multivariate Gaussian distribution.

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

We'll show the derivation step-by-step and also add the comment along the way.

/* Based on the given formula for S_{N+1}^{-1} */

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

/* $\Phi_{1:N+1}^T \Phi_{1:N+1}$ could also be written as $\sum_{n=1}^{N+1} \phi(x_n)\phi(x_n)^T$, where $\phi(x_n)$ is a mx1 vector */

$$= \alpha I_M + \beta \sum_{n=1}^{N+1} \phi(x_n)\phi(x_n)^T$$

$$= \alpha I_M + \beta \sum_{n=1}^{N} \phi(x_n)\phi(x_n)^T + \phi(x_{N+1})\phi(x_{N+1})^T$$

/* The summation of the first two items is equal to S_N^{-1} */

$$= S_N^{-1} + \phi(x_{N+1})\phi(x_{N+1})^T$$

/* x_* is a new test point, so $x_* = x_{N+1}$ */

$$= S_N^{-1} + \phi(x_*)\phi(x_*)^T$$

So, the vector v, in this case, is $\phi(x_{*)}$.

3b: Problem Statement

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

$$(A + vv^T)^{-1} = A^{-1} - \frac{(A^{-1}v)(v^T A^{-1})}{1 + v^T A^{-1} v} \tag{3}$$

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

First, we'll show the left-hand side will be simplified to S_{N+1}:

/* Plug in $A = S_N^{-1}$ and use the result from 3a */

$$(A + vv^T)^{-1} = (S_N^{-1} + vv^T)^{-1} = (S_{N+1}^{-1})^{-1} = S_{N+1} \tag{3b.1}$$

Next, we'll simplify the right hand side of the equation:

/* Plug in $A = S_N^{-1}$ and use the result from 3a */

$$A^{-1} - \frac{(A^{-1}v)(v^T A^{-1})}{1 + v^T A^{-1} v} = S_N - \frac{(S_N \phi(x_*)(\phi(x_*)^T S_N)}{1 + \phi(x_*)^T S_N \phi(x_*)} \tag{3b.2}$$

Combing (3b.1) and (3b.2) together, we have:

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

3c: Problem Statement

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

3c: Solution

First, we write down $\sigma_{N+1}^2(x_*)$ base on the given formula:

$$\sigma_{N+1}^2(x_*) = \beta^{-1} + \phi(x_*)^T S_{N+1} \phi(x_*) \tag{3c.1}$$

Second, we write down $\sigma_N^2(x_*)$ base on the given formula:

$$\sigma_N^2(x_*) = \beta^{-1} + \phi(x_*)^T S_N \phi(x_*) \tag{3c.2}$$

Then, we subtract (3c.2) from (3c.1):

$$\sigma_{N+1}^2(x_*) - \sigma_N^2(x_*) = \phi(x_*)^T (S_{N+1} - S_N) \phi(x_*)$$

3d: Problem Statement

Finally, plug your result from 3b defining S_{N+1} into 3c, plus the fact that S_N must be positive definite, to show that:

$$\sigma^2_{N+1}(x_*) \leq \sigma^2_N(x_*) \tag{4}$$

This would prove that the predictive variance *cannot increase* with each additional data point. In other words, we will never be ”less certain” about a prediction we make if we gather more data.

3d: Solution

Plugging the result from 3b into 3c, we get:

$$\sigma^2_{N+1}(x_*) - \sigma^2_N(x_*) = \phi(x_*)^T \ (-\frac{(S_N\phi(x_*)(\phi(x_*)^T S_N)}{1 + \phi(x_*)^T S_N \phi(x_*)}) \ \phi(x_*)$$

/* The denominator is a positive number since S_N is pos definite (i.e. $\phi(x_*)^T S_N \phi(x_*) \geq 0$) and we always have a constant 1 there */

$$= -\frac{(\phi(x_*)^T \cdot S_N \phi(x_*))(\phi(x_*)^T S_N \cdot \phi(x_*))}{1 + \phi(x_*)^T S_N \phi(x_*)}$$

/* Rewrite the denominator(i.e. $1 + \phi(x_*)^T S_N \phi(x_*)$) as a constant $c > 0$. */

$$= -\frac{(\phi(x_*)^T \cdot S_N \phi(x_*))(\phi(x_*)^T S_N \cdot \phi(x_*))}{c} \tag{3d1}$$

/* Since S_N is pos definite, we know $\phi(x_*)^T S_N \cdot \phi(x_*) \geq 0$ (i.e. it equals to 0 iff $\phi(x_*)$ is a zero vector), the numerator (i.e. $(\phi(x_*)^T \cdot S_N \phi(x_*))(\phi(x_*)^T S_N \cdot \phi(x_*))$) is greater or equal to 0. Thus, the whole fraction is less or equal to zero. */

$$<= 0$$

Thus, we’ve shown that $\sigma^2_{N+1}(x_*) - \sigma^2_N(x_*) \leq 0$.