Student Name: Guangyu Lin

Collaboration Statement:

Total hours spent: 72 hours

I discussed ideas with these individuals:

- Mingyang Wu
- Office hour with TA and Professor
- ...

I consulted the following resources:

- Bishop's textbook
- TODO
- ...

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

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

1a: Solution

$\mathbb{E}[\hat{\sigma}^2(x_1, \dots x_N)]$ substitute the definition given by problem

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

$= \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}(x_n^2 - 2x_n\mu_t + \mu_t^2)]$ then we can calculate the expectation to each term and since μ_t is a known parameter, we can regard it as a constant so we can move the μ_t outside of the expectation due to $\mathbb{E}[x + b] = \mathbb{E}[x] + b$ when b is a constant. On the other hand, the $\frac{1}{N}$ canceled with the sum of N so we can just write μ_t^2 at the outside

$= \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}(x_n^2 - 2x_n\mu_t)] + \mu_t^2$ now we can do the expectation to each term

$= \frac{1}{N}\sum_{n=1}^{N}(\mathbb{E}[x_n^2] - 2\mathbb{E}[\mu_t x_n]) + \mu_t^2$ since μ_t is a constant we can move the μ_t outside from the second term so $2\mathbb{E}[\mu_t x_n]$ can be written as $2\mu_t\mathbb{E}[x_n]$

$=\frac{1}{N}\sum_{n=1}^{N}(\mathbb{E}[x_n^2] - 2\mu_t\mathbb{E}[x_n]) + \mu_t^2$ then by using the property of Gaussian distribution since our variable is drawn from Gaussian model we can find $\mathbb{E}[x_n^2]$ is $\mu_t^2 + \sigma_t^2$ and $\mathbb{E}[x_n]$ is μ_t

$= \frac{1}{N}\sum_{n=1}^{N}(\mu_t^2 + \sigma_t^2 - 2\mu_t^2) + \mu_t^2$ now we can cancel the outside sum with $\frac{1}{N}$

$= \mu_t^2 + \sigma_t^2 - 2\mu_t^2 + \mu_t^2$ simplify this equation

$= \sigma_t^2$

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 is unbiased. From the lecture, we derived the σ_{ml} as $\frac{N-1}{N}\sigma_t^2$. And during that derivation process, σ is an unknown parameter and we only have μ_{ml} which is calculated by the dataset that we collected as $\mu_{ml} = \frac{1}{N}\sum_{n=1}^{N} x_n$. We are not sure about where we drawn these variables so μ_{ml} might be overestimated or underestimated. But for 1a, we have μ_{true} and σ_{true} which are known parameter and using them to generate our variables. Therefore, we derived the expectation of $\hat{\sigma}^2$ is unbiased and based on our result $\frac{N-1}{N}\sigma_t^2$ less than σ_t^2.

2a: Problem Statement

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

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

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

$\log p(x) = \text{const} - \frac{1}{2}(x-\mu)^T S(x-\mu)$ we can write $(x-\mu)^T$ as $(x^T - \mu^T)$

=const $- \frac{1}{2}[(x^T - \mu^T)S(x-\mu)]$ we can expand the $(x^T - \mu^T)S(x-\mu)$

= const $- \frac{1}{2}[x^T Sx - xTS\mu - \mu^T Sx + \mu^T S\mu]$ since S is a symmetric and positive definite matrix $x^T S\mu$ and $\mu^T Sx$ are the same thing and can be combined as one term

= const $-\frac{1}{2}[x^T Sx - 2\mu^T Sx + \mu^T S\mu]$
then we can derive the function $\log p(x) = \text{c} - \frac{1}{2}x^T Ax + b^T x$

$\log p(x) = \text{c} - \frac{1}{2}x^T Ax + b^T x$ we can manipulate $x^T Ax + b^T$ and write it into paranthesis

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

now we can compare this equation and the previous equation that we derived and do the pattern matches

we can find $x^T Ax$ corresponds with $x^T Sx$ and $2b^T x$ corresponds with $2\mu^T Sx$ and since $\mu^T S\mu$ is a constant, we can ignore it and move it to constant part. And now we can say $A := S$ and $b^T := \mu^T S$ in other words $b := \mu S^T$ As long as we know S and μ, we can write the function as $\log p(x) = \text{c} - \frac{1}{2}x^T Ax + b^T x$ form.

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

by definition of $S_N^{-}1$ given by the problem we can write the expression of $S_{N+1}^{-}1$

$S_{N+1}^{-}1 = \alpha I_M + \beta \Phi_{1:N+1}^T \Phi_{1:N+1}$ by the definiction of $\Phi_{1:N}$ we can write $\Phi_{1:N+1}^T \Phi_{1:N+1}$ as $\sum_n^{N+1} \phi(x_n)^T \phi(x_n)$ since the dimension of ϕ is $1 * M$

= $\alpha I_M + \beta \sum_n^{N+1} [\phi(x_n)^T \phi(x_n)]$ now we can rewrite $\beta \sum_n^{N+1} [\phi(x_n)^T \phi(x_n)]$ as $\beta \sum_n^N [\phi(x_n)^T \phi(x_n)] + \beta \phi(x_{n+1})^T \phi(x_{n+1})$

= $\alpha I_M + \beta \sum_n^N [\phi(x_n)^T \phi(x_n)] + \beta \phi(x_{n+1})^T \phi(x_{n+1})$now we can find the first term $\alpha I_M + \beta \sum_n^N [\phi(x_n)^T \phi(x_n)]$ is exactly the definition of $S_N^{-}1$ so we can simplify the equation

= $S_N^-1 + \beta\phi(x_{n+1})^T\phi(x_{n+1})$ now we can do the pattern matches and we can find $\beta\phi(x_{n+1})^T\phi(x_{n+1})$ can be written as $\sqrt{\beta}\phi(x_{n+1})^T * \sqrt{\beta}\phi(x_{n+1})$

by comparing with vv^T we can say $v \coloneqq \sqrt{\beta}\phi(x_{n+1})^T$ and $v^T \coloneqq \sqrt{\beta}\phi(x_{n+1})$

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

by substituting $A = S_N^{-1}$ we can derive the equation that given by problem as

$(S_N^-1 + vv^T)^-1 = S_N - \frac{(S_N v)(v^T S_N)}{1+v^T S_N v}$
by 3a, we know $S_{N+1}^-1 = S_N^{-1} + vv^T$ therefore we can written the previous equation as
$S_{N+1}^-1 = S_N - \frac{(S_N v)(v^T S_N)}{1+v^T S_N v}$

3c: Problem Statement

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

3c: Solution

$\sigma^2_{N+1}(x_*) - \sigma^2_N(x_*)$ substitute with the definition that give by problem

= $[\beta^-1 + \phi(x_*)^T S_{N+1}\phi(x_*)] - [\beta^-1 + \phi(x_*)^T S_N\phi(x_*)]$ we can cancel the β^-1 term

= $\phi(x_*)^T S_{N+1}\phi(x_*) - \phi(x_*)^T S_N\phi(x_*)$ now we can factorize the equation by matrix multiplication law

=$\phi(x_*)^T(S_{N+1}\phi(x_*) - S_N\phi(x_*))$ then we can do the same factorization towards $\phi(x_*)$ by matrix multiplication law

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

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

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

When $\sigma^2_{N+1}(x_*) \leq \sigma^2_N(x_*)$ it must satisfy $\phi(x_*)^T\left[S_{N+1} - S_N\right]\phi(x_*) \leq 0$ then we can substitute $S_{N+1} - S_N$ with its definition given by the problem and we can get $\phi(x_*)^T\left[S_N - \frac{S_N v v^T S_N}{1+v^T S_N V} - S_N\right]\phi(x_*)$ S_N will be canceled by each other
then we will only have $\phi(x_*)^T\left[-\frac{S_N v v^T S_N}{1+v^T S_N V}\right]\phi(x_*)$
and it can be written as $[-\frac{\phi(x_*)^T S_N v v^T S_N \phi(x_*)}{1+v^T S_N V}]$
now we can split the numerator as two terms, one is $\phi(x_*)^T S_N v$ and another one is $v^T S_N \phi(x_*)$
Since S_N is the covariance matrix and it is def positive and symmetric, we can regard these two terms are same. Therefore, the numerator can be regard as $(\phi(x_*)^T S_N v)^2$. And the quadratic form will always bigger or equal to 0. Then for the numerator, since we know S_N is a covariance matrix then by its property we know $v^T S_N v$ will always greater than or equal to 0. Therefore we can say $\frac{\phi(x_*)^T S_N v v^T S_N \phi(x_*)}{1+v^T S_N V}$ must greater than or equal to 0. And since there is a negative sign before this term, we can say $-\frac{\phi(x_*)^T S_N v v^T S_N \phi(x_*)}{1+v^T S_N V}$ must less than or equal to 0.

Therefore $\sigma^2_{N+1}(x_*) \leq \sigma^2_N(x_*)$ is true.