Linear MMSE

1. Linear MMSE

데이터 $x[0], x[1], \ldots, x[N-1]$이 주어졌을 때, 스칼라 파라미터 θ를 추정한다고 하자. linear이기 때문에 추정값은 다음과 같이 선형 결합으로 표현될 것이다.

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N$$

이를 구하기 위해 다음과 같은 Bmse 값을 최소화해야 한다.

$$Bmse(\hat{\theta}) = E[(\theta - \hat{\theta})^2]$$

다만 구하고자 하는 추정값이 선형이 아닌 경우, Linear MMSE를 항상 최적의 추정값이라고 볼 수는 없다. 예를 들어 단일 관측 데이터 $x[0]$에 대해 $x[0]^2$를 추정한다고 가정해 보자. 그렇다면 다음의 값이 최적의 추정값일 것이다.

$$\hat{\theta} = x[0]^2$$

하지만 LMMSE에서 찾는 추정값은 다음의 형태이다.

$$\hat{\theta} = a_0 x[0] + a_1$$

Bmse를 구해보면 다음과 같다.

$$Bmse(\hat{\theta}) = E[(\theta - \hat{\theta})^2] = E[(\theta - a_0 x[0] - a_1)^2]$$

위 Bmse의 최소를 구하기 위해 각 parameter로 미분하면 다음과 같다.

$$\frac{\partial}{\partial a_0} Bmse(\hat{\theta}) = E[(\theta - a_0 x[0] - a_1) x[0]] = 0$$

$$a_0 E(x[0]^2) = 0$$

$$a_0 = 0$$

$$\frac{\partial}{\partial a_1} Bmse(\hat{\theta}) = E[(\theta - a_0 x[0] - a_1)] = 0$$

$$a_1 = E(\theta) = E(x[0]^2) = \sigma^2$$

따라서 구하고자 하는 LMMSE와 최소 Bmse값은 다음과 같다.

$$LMMSE\,\hat{\theta} = \sigma^2$$

$$Bmse(\hat{\theta}) = 2\sigma^4$$

위 값은 $\hat{\theta} = x[0]^2$일 때에 비해 Bmse가 큰 것으로 보아 최적은 아닌 것으로 보인다. 따라서 LMMSE는 항상 최적이 아닐 수 있다.

일반적인 경우에 LMMSE를 찾아보자. Bmse를 미분해보면 다음과 같다.

$$\frac{\partial}{\partial a_N} E[(\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N)^2] = -2E[\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N] = 0$$

$$a_N = E(\theta) - \sum_{n=0}^{N-1} a_n E(x[n])$$

위 값을 다시 Bmse 식에 대입하고 미분하면 다음과 같다.

$$Bmse(\hat{\theta}) = E[\sum_{n=0}^{N-1} a_n (x[n] - E(x[n])) - (\theta - E(\theta))]^2$$

$$
\begin{aligned}
Bmse(\hat{\theta}) &= E[a^T(x-E(x))-(\theta-E(\theta))]^2 \\
&= E[a^T(x-E(x))(x-E(x))^T a] - E[a^T(x-E(x))(\theta-E(\theta))] \\
&\quad - E[(\theta-E(\theta))(x-E(x))^T a] + E[(\theta-E(\theta))^2] \\
&= a^T C_{xx} a - a^T C_{x\theta} - C_{\theta x} a + C_{\theta\theta}
\end{aligned}
$$

$$\frac{\partial Bmse(\hat{\theta})}{\partial a} = 2C_{xx}a - 2C_{x\theta} = 0$$

$$a = C_{xx}^{-1} C_{x\theta}$$

$$\hat{\theta} = a^T x + a_N = C_{x\theta}^T C_{xx}^{-1} x + E(\theta) - C_{x\theta}^T C_{xx}^{-1} E(x)$$

따라서 구하고자 하는 LMMSE는 다음과 같다.

$$\hat{\theta} = E(\theta) + C_{\theta x} C_{xx}^{-1}(x - E(x))$$

또한 이 때의 Minimum Bmse는 다음과 같다.

$$
\begin{aligned}
Bmse(\hat{\theta}) &= a^T C_{xx} a - a^T C_{x\theta} - C_{\theta x} a + C_{\theta\theta} \\
&= C_{\theta\theta} - C_{\theta x} C_{xx}^{-1} C_{x\theta}
\end{aligned}
$$

예시로 다음의 문제를 풀어보자.

다음과 같은 데이터가 주어졌다.

$$x[n] = A + w[n],\ n = 0, 1, \ldots, N-1 \text{ where } A \sim u[-A_0, A_0],\ w[n] \sim N(0, \sigma^2)$$

그렇다면 LMMSE는 다음과 같이 계산할 수 있다.

$$E(x) = 0,\ E(A) = 0$$

$$
\begin{aligned}
C_{xx} &= E(xx^T) \\
&= E[(A1 + w)(A1 + w)^T] \\
&= E(A^2)11^T + \sigma^2 I \\
&= \sigma_A^2 11^T + \sigma^2 I
\end{aligned}
$$

$$
\begin{aligned}
C_{\theta x} &= E(Ax^T) \\
&= E(A(A1 + w)^T) \\
&= E(A^2)1^T \\
&= \sigma_A^2 1^T
\end{aligned}
$$

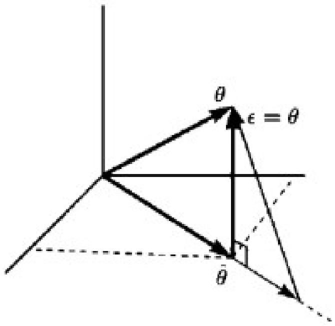
$$
\begin{aligned}
\hat{A} &= C_{\theta x} C_{xx}^{-1} x \\
&= \sigma_A^2 1^T (\sigma_A^2 11^T + \sigma^2 I)^{-1} x \\
&= \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} \\
&= \frac{\frac{A_0^2}{3}}{\frac{A_0^2}{3} + \frac{\sigma^2}{N}} \bar{x}
\end{aligned}
$$

2. Geometrical Interpretations

Vector Space 내의 두 Random Variables X, Y가 있을 때, 두 변수 간 내적은 E(XY)와 같으며 E(XY)가 0이면 두 변수는 orthogonal하다.

LMMSE $\hat{\theta}=\sum_{n=0}^{N-1} a_n x[n]$은 다음 Bmse값을 최소화 한다.

$$Bmse(\hat{\theta})=E[(\theta-\hat{\theta})^2]=E[(\theta-\sum_{n=0}^{N-1} a_n x[n])^2]=\|\theta-\sum_{n=0}^{N-1} a_n x[n]\|^2$$



우선 LMMSE는 $x[0],x[1],x[2],...,x[N-1]$이 span하는 subspace인 S위의 벡터이다.
따라서 Bmse가 최소가 되기 위해선 위 그림처럼 LMMSE가 θ를 S위로 내린 projection 이어야 한다. 따라서 $\theta-\hat{\theta}=\epsilon\perp S$이며 $\epsilon\perp x[0],x[1],...,x[N-1]$이다.
따라서 $E[(\theta-\hat{\theta})x[n]]=0$ for $n=0,1,...,N-1$이 만족한다.
이에 따라 다음과 같이 전개할 수 있다.

$$E[(\theta-\sum_{m=0}^{N-1} a_m x[m])x[n]]=0$$

$$\sum_{m=0}^{N-1} a_m E(x[m]x[n])=E(\theta x[n])$$

$$\begin{bmatrix} E(x^2[0]) & E(x[0]x[1]) & \dots & E(x[0]x[N-1]) \\ E(x[1]x[0]) & E(x^2[1]) & \dots & E(x[1]x[N-1]) \\ \dots & \dots & \dots & \dots \\ E(x[N-1]x[0]) & E(x[N-1]x[1]) & \dots & E(x^2[N-1]) \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_{N-1}\end{bmatrix}=\begin{bmatrix} E(\theta x[0]) \\ E(\theta x[1]) \\ \dots \\ E(\theta x[N-1])\end{bmatrix}$$

$$\begin{aligned} C_{xx}a &= C_{x\theta} \\ a &= C_{xx}^{-1}C_{x\theta} \\ \hat{\theta} &= a^T x \\ &= C_{x\theta}^T C_{xx}^{-1}x = C_{\theta x}C_{xx}^{-1}x \end{aligned}$$

$$\begin{aligned} Bmse(\hat{\theta}) &= E[(\theta-\sum_{n=0}^{N-1} a_n x[n])(\theta-\sum_{m=0}^{N-1} a_m x[m])] \\ &= E[(\theta-\sum_{n=0}^{N-1} a_n x[n])\theta]-E[(\theta-\sum_{n=0}^{N-1} a_n x[n])\sum_{m=0}^{N-1} a_m x[m]] \\ &= E(\theta^2)-\sum_{n=0}^{N-1} a_n E(x[n]\theta)-\sum_{m=0}^{N-1} a_m E[(\theta-\sum_{n=0}^{N-1} a_n x[n])x[m]] \\ &= C_{\theta\theta}-a^T C_{x\theta} \\ &= C_{\theta\theta}-C_{x\theta}^T C_{xx}^{-1}C_{x\theta} \\ &= C_{\theta\theta}-C_{\theta x}C_{xx}^{-1}C_{x\theta} \end{aligned}$$

3. Vector LMMSE

p개의 추정값이 모인 vector estimator가 있을 때 각각은 다음과 같이 표현가능하다.

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] + a_{iN} \quad \text{for } i = 1, 2, ..., p$$

그리고 이는 앞선 풀이에 따라 다음과 같이 표현 가능하다.

$$\hat{\theta}_i = E(\theta_i) + C_{\theta_i x} C_{xx}^{-1} (x - E(x))$$

$$Bmse(\hat{\theta}_i) = C_{\theta_i \theta_i} - C_{\theta_i x} C_{xx}^{-1} C_{x \theta_i}$$

이를 벡터로 표현하면 다음과 같다.

$$\begin{aligned} \hat{\theta} &= \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \cdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} C_{\theta_1 x} C_{xx}^{-1} (x - E(x)) \\ C_{\theta_2 x} C_{xx}^{-1} (x - E(x)) \\ \cdots \\ C_{\theta_p x} C_{xx}^{-1} (x - E(x)) \end{bmatrix} \\ &= \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \cdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} C_{\theta_1 x} \\ C_{\theta_2 x} \\ \cdots \\ C_{\theta_p x} \end{bmatrix} C_{xx}^{-1} (x - E(x)) \\ &= E(\theta) + C_{\theta x} C_{xx}^{-1} (x - E(x)) \end{aligned}$$

$$\begin{aligned} M_{\hat{\theta}} &= E[(\theta - \hat{\theta})(\theta - \hat{\theta})^T] \\ &= C_{\theta\theta} - C_{\theta x} C_{xx}^{-1} C_{x\theta} \end{aligned}$$

$$Bmse(\hat{\theta}_i) = [M_{\hat{\theta}}]_{ii}$$

4. Properties of the LMMSE

1) linear transformation에 대해 보존된다.

$$\alpha = A\theta + b \rightarrow \hat{\alpha} = A\hat{\theta} + b$$

2) 두 LMMSE 추정값의 선형 결합에 대해 성립한다.

$$\alpha = \theta_1 + \theta_2 \rightarrow \hat{\alpha} = \hat{\theta}_1 + \hat{\theta}_2$$

3) MMSE가 linear하면 LMMSE는 MMSE의 선형 가우시안 모델이다.

Theorem 12.1 (Bayesian Gauss-Markov Theorem) *If the data are described by the Bayesian linear model form*

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \tag{12.25}$$

where $\mathbf{x}$ is an $N \times 1$ data vector, $\mathbf{H}$ is a known $N \times p$ observation matrix, $\boldsymbol{\theta}$ is a $p \times 1$ random vector of parameters whose realization is to be estimated and has mean $E(\boldsymbol{\theta})$ and covariance matrix $\mathbf{C}_{\theta\theta}$, and $\mathbf{w}$ is an $N \times 1$ random vector with zero mean and covariance matrix $\mathbf{C}_w$ and is uncorrelated with $\boldsymbol{\theta}$ (the joint PDF $p(\mathbf{w}, \boldsymbol{\theta})$ is otherwise arbitrary), then the LMMSE estimator of $\boldsymbol{\theta}$ is

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}) + \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})) & (12.26) \\ &= E(\boldsymbol{\theta}) + (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_w^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})). & (12.27) \end{aligned}$$

The performance of the estimator is measured by the error $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ whose mean is zero and whose covariance matrix is

$$\begin{aligned} \mathbf{C}_\epsilon &= E_{x,\theta}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_{\theta\theta} & (12.28) \\ &= (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}. & (12.29) \end{aligned}$$

The error covariance matrix is also the minimum MSE matrix $\mathbf{M}_{\hat{\theta}}$ whose diagonal elements yield the minimum Bayesian MSE

$$\begin{aligned} [\mathbf{M}_{\hat{\theta}}]_{ii} &= [\mathbf{C}_\epsilon]_{ii} \\ &= \mathrm{Bmse}(\hat{\theta}_i). & (12.30) \end{aligned}$$