Linear MMSE for Random Vectors

Jiman Hwang

This document is for self-study only.
Source: [PN]

1 Statement

Suppose that we would like to have an estimator for the random vector $\mathbf{X}$ in the form of

$$\hat{\mathbf{X}}_L = \mathbf{AY} + \mathbf{b}$$

then

$$\hat{\mathbf{X}}_L = \mathbf{C_{XY}}{\mathbf{C_Y}}^{-1}\left(\mathbf{Y} - \mathrm{E}[\mathbf{Y}]\right) + \mathrm{E}[\mathbf{X}].$$

where $\mathbf{C_Y}$ is the covariance matrix of $\mathbf{Y}$,

$$\mathbf{C_Y} = \mathrm{E}[(\mathbf{Y} - E\mathbf{Y})(\mathbf{Y} - E\mathbf{Y})^T],$$

and $\mathbf{C_{XY}}$ is the cross covariance matrix of $\mathbf{X}$ and $\mathbf{X}$,

$$\mathbf{C_{XY}} = \mathrm{E}[(\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T].$$

2 Derivation

Here's basic setup.

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \hat{\mathbf{X}}_L = \mathbf{AY} + \mathbf{b} \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_m^T \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \tag{1}$$

The error $\tilde{\mathbf{X}}$ is defined as

$$\tilde{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{X}}_L$$

and we should minimize Mean Squared Error(MSE), i.e,

$$\min_{\hat{\mathbf{X}}_L} \mathrm{E}(\|\tilde{\mathbf{X}}\|^2) = \min_{\hat{\mathbf{X}}_L} \mathrm{E}(\|\mathbf{X} - \hat{\mathbf{X}}_L\|^2) = \min_{\mathbf{A},\mathbf{b}} \mathrm{E}(\|\mathbf{X} - \mathbf{AY} - \mathbf{b}\|^2)$$

Let

$$h(\mathbf{A}, \mathbf{b}) = \mathrm{E}(\|\mathbf{X} - \mathbf{AY} - \mathbf{b}\|^2)$$

According to (1),

$$h(\mathbf{A}, \mathbf{b}) = \mathrm{E}\left[\sum_{k=1}^{m}\left(X_k - \mathbf{v}_k^T\mathbf{Y} - b_k\right)^2\right] \tag{2}$$

Let the critical point $(\mathbf{A}_L, \mathbf{b}_L)$ of h be

$$\mathbf{A}_L = \begin{bmatrix} \mathbf{v}_{L1}^T \\ \vdots \\ \mathbf{v}_{Lm}^T \end{bmatrix} = \begin{bmatrix} a_{L11} & \cdots & a_{L1n} \\ \vdots & & \vdots \\ a_{Lm1} & \cdots & a_{Lmn} \end{bmatrix}, \mathbf{b}_L = \begin{bmatrix} b_{L1} \\ \vdots \\ b_{Lm} \end{bmatrix}$$

To find minimum, differentiate (2) with respect to each element of $\mathbf{A}$.

$$\begin{aligned} \frac{\partial h}{\partial a_{ij}} =& \mathrm{E}\left[2\left(X_i - \mathbf{v}_i^T\mathbf{Y} - b_i\right)(-Y_j)\right] \\ =& \mathrm{E}\left[-2\left(X_iY_j - \left(\mathbf{v}_i^T\mathbf{Y}\right)Y_j - b_iY_j\right)\right] \end{aligned}$$

Since $\mathbf{v}_i^T\mathbf{Y} = \sum_{k=1}^n a_{ik}Y_k$,

$$\frac{\partial h}{\partial a_{ij}} = \mathrm{E}\left[-2\left(X_iY_j - \sum_{k=1}^n a_{ik}Y_kY_j - b_iY_j\right)\right]$$

At critical point,

$$\begin{aligned}
\left.\frac{\partial h}{\partial a_{ij}}\right|_{\mathbf{A}_L,\mathbf{b}_L} =& \mathrm{E}\left[-2\left(X_iY_j - \sum_{k=1}^n a_{Lik}Y_kY_j - b_{Li}Y_j\right)\right] = 0 \\
&\Rightarrow \mathrm{E}\left[X_iY_j - \sum_{k=1}^n a_{Lik}Y_kY_j - b_{Li}Y_j\right] = 0 \\
&\Rightarrow \mathrm{E}\left[X_iY_j\right] - \sum_{k=1}^n a_{Lik}\mathrm{E}\left[Y_kY_j\right] - b_{Li}\mathrm{E}\left[Y_j\right] = 0 \\
&\Rightarrow \sum_{k=1}^n a_{Lik}\mathrm{E}\left[Y_kY_j\right] + b_{Li}\mathrm{E}\left[Y_j\right] = \mathrm{E}\left[X_iY_j\right] \\
&\Rightarrow \mathbf{v}_{Li}^T\begin{bmatrix}\mathrm{E}\left[Y_1Y_j\right]\\ \vdots \\ \mathrm{E}\left[Y_nY_j\right]\end{bmatrix} + b_{Li}\mathrm{E}\left[Y_j\right] = \mathrm{E}\left[X_iY_j\right]
\end{aligned}$$

Setting $i = 1, \cdots, m$ and assemble(?) them in rows.

$$\begin{aligned}
&\begin{bmatrix}\mathbf{v}_{L1}^T\\ \vdots \\ \mathbf{v}_{Lm}^T\end{bmatrix}\begin{bmatrix}\mathrm{E}\left[Y_1Y_j\right]\\ \vdots \\ \mathrm{E}\left[Y_nY_j\right]\end{bmatrix} + \begin{bmatrix}b_{L1}\\ \vdots \\ b_{Lm}\end{bmatrix}\mathrm{E}\left[Y_j\right] = \begin{bmatrix}\mathrm{E}\left[X_1Y_j\right]\\ \vdots \\ \mathrm{E}\left[X_mY_j\right]\end{bmatrix} \\
\Rightarrow \mathbf{A}_L&\begin{bmatrix}\mathrm{E}\left[Y_1Y_j\right]\\ \vdots \\ \mathrm{E}\left[Y_nY_j\right]\end{bmatrix} + \mathbf{b}_L\mathrm{E}\left[Y_j\right] = \begin{bmatrix}\mathrm{E}\left[X_1Y_j\right]\\ \vdots \\ \mathrm{E}\left[X_mY_j\right]\end{bmatrix}
\end{aligned}$$

Setting $j = 1, \cdots, n$ and assemble(?) them in columns.

$$\mathbf{A}_L\begin{bmatrix}\mathrm{E}\left[Y_1Y_1\right] & \cdots & \mathrm{E}\left[Y_1Y_n\right]\\ \vdots & & \vdots \\ \mathrm{E}\left[Y_nY_1\right] & \cdots & \mathrm{E}\left[Y_nY_n\right]\end{bmatrix} + \mathbf{b}_L\begin{bmatrix}\mathrm{E}\left[Y_1\right] & \cdots & \mathrm{E}\left[Y_n\right]\end{bmatrix} = \begin{bmatrix}\mathrm{E}\left[X_1Y_1\right] & \cdots & \mathrm{E}\left[X_1Y_n\right]\\ \vdots & & \vdots \\ \mathrm{E}\left[X_mY_1\right] & \cdots & \mathrm{E}\left[X_mY_n\right]\end{bmatrix}$$

Cleaning up with vectors,

$$\mathbf{A}_L\mathrm{E}\left[\mathbf{Y}\mathbf{Y}^T\right] + \mathbf{b}_L\mathrm{E}\left[\mathbf{Y}^T\right] = \mathrm{E}\left[\mathbf{X}\mathbf{Y}^T\right] \tag{3}$$

Differentiating (2) breeds another equation.

$$\frac{\partial h}{\partial a_{ij}} = \mathrm{E}\left[-2\left(X_i - \mathbf{v}_i^T\mathbf{Y} - b_i\right)\right]$$

Again at critical point,

$$\begin{aligned}
\left.\frac{\partial h}{\partial a_{ij}}\right|_{\mathbf{A}_L,\mathbf{b}_L} =& \mathrm{E}\left[-2\left(X_i - \mathbf{v}_{Li}^T\mathbf{Y} - b_{Li}\right)\right] = 0 \\
&\Rightarrow \mathrm{E}\left[X_i - \mathbf{v}_{Li}^T\mathbf{Y} - b_{Li}\right] = 0 \\
&\Rightarrow \mathrm{E}\left[X_i\right] - \sum_{k=1}^n a_{Lik}\mathrm{E}\left[Y_k\right] - b_{Li} = 0 \\
&\Rightarrow \sum_{k=1}^n a_{Lik}\mathrm{E}\left[Y_k\right] + b_{Li} = \mathrm{E}\left[X_i\right] \\
&\Rightarrow \mathbf{v}_{Li}^T\begin{bmatrix}\mathrm{E}\left[Y_1\right]\\ \vdots \\ \mathrm{E}\left[Y_n\right]\end{bmatrix} + b_{Li} = \mathrm{E}\left[X_i\right]
\end{aligned}$$

Setting $i = 1, \cdots, m$ and assemble(?) them in rows.

$$\begin{bmatrix} \mathbf{v}_{L1}^T \\ \vdots \\ \mathbf{v}_{Lm}^T \end{bmatrix} \begin{bmatrix} \mathrm{E}[Y_1] \\ \vdots \\ \mathrm{E}[Y_n] \end{bmatrix} + \begin{bmatrix} b_{L1} \\ \vdots \\ b_{Lm} \end{bmatrix} = \begin{bmatrix} \mathrm{E}[X_1] \\ \vdots \\ \mathrm{E}[X_m] \end{bmatrix}$$

Cleaning up with vectors,

$$\mathbf{A}_L \mathrm{E}[\mathbf{Y}] + \mathbf{b}_L = \mathrm{E}[\mathbf{X}]$$

or

$$\mathbf{b}_L = \mathrm{E}[\mathbf{X}] - \mathbf{A}_L \mathrm{E}[\mathbf{Y}] \tag{4}$$

Putting this into (3),

$$\begin{aligned} & \mathbf{A}_L \mathrm{E}\left[\mathbf{Y}\mathbf{Y}^T\right] + \left(\mathrm{E}[\mathbf{X}] - \mathbf{A}_L \mathrm{E}[\mathbf{Y}]\right) \mathrm{E}\left[\mathbf{Y}^T\right] = \mathrm{E}\left[\mathbf{X}\mathbf{Y}^T\right] \\ \Rightarrow\ & \mathbf{A}_L \left(\mathrm{E}\left[\mathbf{Y}\mathbf{Y}^T\right] - \mathrm{E}[\mathbf{Y}]\, \mathrm{E}\left[\mathbf{Y}^T\right]\right) = \mathrm{E}\left[\mathbf{X}\mathbf{Y}^T\right] - \mathrm{E}[\mathbf{X}]\, \mathrm{E}\left[\mathbf{Y}^T\right] \\ \Rightarrow\ & \mathbf{A}_L \mathbf{C}_{\mathbf{Y}} = \mathbf{C}_{\mathbf{XY}} \end{aligned}$$

So we get

$$\mathbf{A}_L = \mathbf{C}_{\mathbf{XY}} \mathbf{C}_{\mathbf{Y}}^{-1} \tag{5}$$

Putting this into (4),

$$\mathbf{b}_L = \mathrm{E}[\mathbf{X}] - \mathbf{C}_{\mathbf{XY}} \mathbf{C}_{\mathbf{Y}}^{-1} \mathrm{E}[\mathbf{Y}]$$

Finally the linear MMSE $\hat{\mathbf{X}}_L$ would be

$$\begin{aligned} \hat{\mathbf{X}}_L = & \mathbf{A}_L \mathbf{Y} + \mathbf{b}_L \\ = & \mathbf{C}_{\mathbf{XY}} \mathbf{C}_{\mathbf{Y}}^{-1} \mathbf{Y} + \mathrm{E}[\mathbf{X}] - \mathbf{C}_{\mathbf{XY}} \mathbf{C}_{\mathbf{Y}}^{-1} \mathrm{E}[\mathbf{Y}] \\ = & \mathbf{C}_{\mathbf{XY}} \mathbf{C}_{\mathbf{Y}}^{-1} \left(\mathbf{Y} - \mathrm{E}[\mathbf{Y}]\right) + \mathrm{E}[\mathbf{X}] \end{aligned}$$

References

[PN] Hossein Pishro-Nik. Mean Squared Error (MSE). `http://www.probabilitycourse.com/chapter9/9_1_7_estimation_for_random_vectors.php`.