LINEAR BAYESIAN ESTIMATORS

Presented by
Dheeraj Kumar (Y5827167)
Varun Khaitan (Y5495)

LINEAR MINIMUM MEAN SQUARE ERROR(LMMSE) ESTIMATION

LMMSE

- Consider the data set {x[0], x[1],…,x[N-1]} or in vector form as **X** = [x[0] x[1] … x[N-1]]T
- Aim is to estimate unknown scalar parameter θ based on the data set.
- This method does not assume any specific form for the joint pdf p(**X**, θ) but only a knowledge of first two moments.
- For a linear estimator, we rely on the correlation between **X** and θ.

LMMSE (Contd…)

- Consider the class of all linear (affine) estimators of the form-

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

… (1)

- We choose the coefficients a_n's so as to minimize the Bayesian MSE

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

… (2)

- To minimize the MSE, differentiate above equation wrt all a_n's and set it to 0.

LMMSE (Contd…)

- Substituting equation (1) in (2) and differentiating wrt a_N and setting the result to 0 yields-

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

- Setting it to zero yields-

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

LMMSE (Contd…)

- So now we need to minimize-

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

- Let **a** = $[a_0\ a_1\ \ldots\ a_{N-1}]^T$, we have-

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

Where C_{XX} is N×N covariance matrix of X, and $C_{\theta X}$ is 1×N cross covariance vector having the property that $C_{\theta X}^T = C_{X\theta}$, and $C_{\theta\theta}$ is the variance of θ

LMMSE (Contd…)

- To minimize Bmse, we set-

$$\frac{\partial \mathrm{Bmse}(\hat{\theta})}{\partial \mathbf{a}} = 2\mathbf{C}_{xx}\mathbf{a} - 2\mathbf{C}_{x\theta} = \mathbf{0}$$

$$\mathbf{a} = \mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}.$$

Which gives,
$$\begin{aligned}\hat{\theta} &= \mathbf{a}^T\mathbf{x} + a_N \\ &= \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}\mathbf{x} + E(\theta) - \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}E(\mathbf{x}) \\ &= E(\theta) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})).\end{aligned}$$

- Minimum Bmse is given by-

$$\begin{aligned}\mathrm{Bmse}(\hat{\theta}) &= \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}\mathbf{C}_{xx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} - \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \\ &\quad - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} + C_{\theta\theta} \\ &= \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} - 2\mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} + C_{\theta\theta} \\ &= C_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}.\end{aligned}$$

GEOMETRICAL INTERPRETATIONS (GI)

GI

- LMMSE estimator admits a geometrical interpretation where "vectors" are random variables.
- This formulation assumes that θ and **x** are zero mean random variables. If they are not we can always define zero mean random variables: θ' = θ – E(θ) ; **x** = **x** – E(**x**)
- Consider estimation of θ' as linear function of **x**.

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

And minimizes $\mathrm{Bmse}(\hat{\theta}) = E\left[(\theta - \hat{\theta})^2\right]$

GI (Contd…)

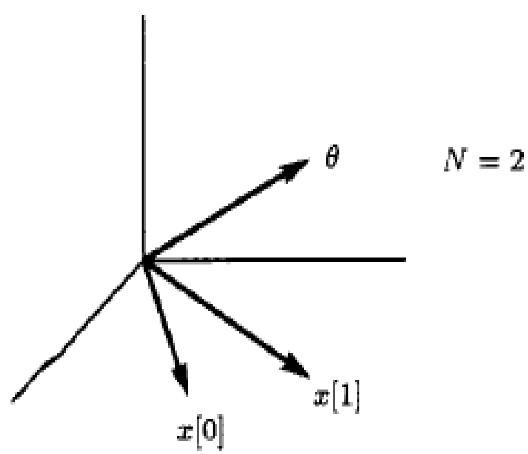
- Consider θ, x[0], x[1],…,x[N-1] as elements in vector space.
- Length of each vector is defined as $||x|| = (E(x^2))^{1/2}$, which is square root of variance.
- Inner product of vectors x and y is : $(x,y) = E(xy)$

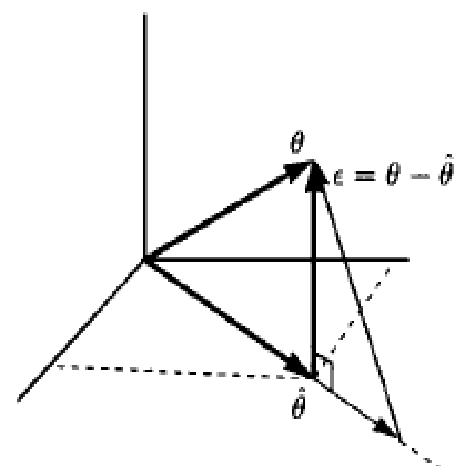
GI (Contd…)

- Bayesian mean square error is given as

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

- Minimization of MSE is equivalent to minimize the square length of the error vector, $\epsilon = \theta - \hat{\theta}$.





Orthogonality principle

- Orthogonality principal/ Projection theorem: While estimating the realization of a random variable by a linear combination of data samples, the optimal estimator is obtained when the error is orthogonal to each data sample.

$$\epsilon \perp x[0], x[1], \ldots, x[N-1]$$

$$E\left[(\theta - \hat{\theta})x[n]\right] = 0 \qquad n = 0, 1, \ldots, N-1.$$

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

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

GI (Contd…)

- These equations can be written in matrix form as:

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

- $\mathbf{C}_{xx}\,\mathbf{a} = \mathbf{C}_{x\theta}$ => $\mathbf{a} = \mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}$
- LMMSE estimate of θ is:

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

GI (Contd…)

- Minimum Bayesian MSE is given by squared length of error vector

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

THANK YOU