Linear LMS Estimation

mardi 19 avril 2022 17:39

Linear LMS Estimation

- To find LMS, need full statistics $f(\theta, x)$
- Only have 1st & 2nd moments of Θ, X: $E[\Theta], E[X], Var(X), Var(\Theta), Cov(\Theta, X)$
  ↳ can compute estimate $\hat{\Theta} = aX + b$ minimizing $MSE = E[(\Theta - \hat{\Theta})^2]$
  ↳ if a found, what is best b?

$$E[(\Theta-\hat{\Theta})^2] = E[(\Theta - aX - b)^2] = E[(Y-b)^2] = E[Y^2] - 2E[Y]\cdot b + b^2 \triangleq g(b)$$

$$\text{Set } \frac{d\,g(b)}{db} = 2b - 2E[Y] = 0 \longrightarrow b = E[Y] = E[\Theta - aX]$$

- Find best a minimizing $E[(\Theta - aX - \overbrace{E[\Theta - aX]}^{b})^2]$

$$g(a) = E[(\Theta - aX - E[\Theta - aX])^2]$$
$$= Var(\Theta - aX) = Var(\Theta) - 2a\,Cov(\Theta, X) + a^2 Var X$$

$$\text{Set } \frac{d\,g(a)}{da} = 2\,Var X\, a - 2Cov(\Theta, X) = 0 \longrightarrow a = \frac{Cov(\Theta, X)}{Var X}$$

$$\therefore \hat{\Theta}_{LLMS} = aX + b = \frac{Cov(\Theta, X)}{Var X} X + E[\Theta] - \frac{Cov(\Theta, X)}{Var X} \cdot EX$$

Corresponding MSE

$$- MSE_{LLMS} = E[(\Theta - \hat{\Theta})^2] = E[(\Theta - E\Theta - \frac{Cov(\Theta, X)}{Var X}(X - EX))^2]$$
$$= \dots = Var\,\Theta - \frac{[Cov(\Theta, X)]^2}{Var X} = Var\,\Theta(1 - \rho_{\Theta,X}^2)$$

Linear LMS vs LMS Estimation

- $\hat{\Theta}_{LLMS} \neq \hat{\Theta}_{LMS}$, $MSE_{LLMS} \geq MSE_{LMS}$
- Special Case: when Θ, X jointly Gaussian, $\hat{\Theta}_{LLMS} = \hat{\Theta}_{LMS}$

Joint Gaussian RV

- RVs = jointly Gaussian if joint pdf has form:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{x,y}^2}} e^{-\frac{1}{2(1-\rho_{x,y}^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho_{x,y}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}$$

↳ pdf is determined $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{x,y}$

- Example 1: $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$, X and Y are independent, $\Rightarrow \rho_{x,y} = 0$
  then X and Y are jointly Gaussian with joint pdf

$$f(x,y) = f(x)f(y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]}$$

- Contours of equal joint pdf = ellipses w/ quadratic eqn:

$$\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2} - 2\rho_{x,y}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} = c \geq 0$$

- Orientation of major axis of ellipse:

$$\theta = \frac{1}{2}\arctan\left(\frac{2\rho_{xy}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}\right)$$

Jointly Gaussian Random Vector

Jointly Gaussian random vector

- Random variables $X_1, X_2, \dots, X_n$ are jointly Gaussian, or the random vector $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ is Gaussian $N(\underline{\mu}, \Sigma)$, if the joint pdf is of the form.

$$\boxed{f(x_1, \dots, x_n) = f(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})},}$$

Jointly Gaussian random vector

- Random variables $X_1, X_2, \cdots, X_n$ are jointly Gaussian, or the random vector $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ is Gaussian $N(\underline{\mu}, \Sigma)$, if the joint pdf is of the form.

$$\boxed{f(x_1, \cdots, x_n) = f(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})}},$$

where $\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\underline{\mu} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$, $\Sigma = \begin{bmatrix} Cov(X_1,X_1), & \cdots & Cov(X_1,X_n) \\ & Cov(X_i, X_j) & \\ Cov(X_n X_1) & \cdots & Cov(X_n, X_n) \end{bmatrix}$ and $\det(\Sigma) > 0$.

Covariance matrix.

Properties of Jointly Gaussian

① Linear Transforms = jointly gaussian
↳ given $m \times n$ full rank matrix A w/ $m \le n$, $x \sim N(\underline{\mu}, \Sigma)$
$\underline{Y} = Ax \sim N(A\underline{\mu}, A\Sigma A^T)$

② Marginals of Jointly Gaussian RVs = Jointly gaussian

③ Conditionals of Jointly Gaussian RVs = JG if:

$$\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix} = N\left(\begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

where $\underline{X}$ is an n-dim vector, $\underline{X}_1$ is a k-dim vector, $\underline{X}_2$ is an (n-k)-dim vector,

then, $\boxed{\underline{X}_2 \mid \{\underline{X}_1 = \underline{x}_1\} \sim N\left(\Sigma_{21}\Sigma_{11}^{-1}(\underline{x}_1 - \underline{\mu}_1) + \underline{\mu}_2, \ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)}$.

Linear LMS vs LMS vs MAP estimator

- If unknown Θ & data x (can be vector?) = JG,
$\therefore \hat{\Theta}_{LLMS} = \hat{\Theta}_{LMS} = \hat{\Theta}_{MAP}$

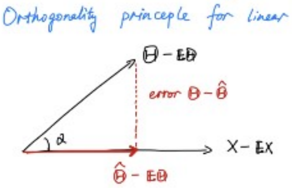
Example: Estimating Gaussian signal in Gaussian noise.

- Signal $\Theta \sim N(0,1)$, $X = \Theta + W$, $W \sim N(0,1)$ indep. of Θ
- $\begin{bmatrix} X \\ \Theta \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} W \\ \Theta \end{bmatrix} \Rightarrow X$ and Θ are jointly Gaussian
- $\Rightarrow \hat{\Theta}_{LLMS} = \hat{\Theta}_{LMS} = \hat{\Theta}_{MAP} = \frac{X}{2}$

Geometric Formulation of Linear Estimation

- Vector Space V → Set of Vectors closed under two operations:
  ↳ Vector Addition: $v_1, v_2 \in V \rightarrow v_1 + v_2 \in V$
  ↳ Scalar Multiplication: if $a \in \mathbb{R}, v \in V, \rightarrow a \cdot v \in V$
- Inner product
  ↳ Commutativity: $u^T v = v^T u$
  ↳ Linearity: $(au+v)^T w = au^T w + v^T w$
  ↳ Nonnegativity: $u^T u \geq 0$ and $u^T u = 0$ iff $u = 0$
- Norm u: $\|u\| = \sqrt{u^T u}$
- $u \perp v$ → orthogonal if $u^T v = 0$

Orthogonality principle for linear LMS estimation.



inner product ⟺ $Cov(\Theta, X)$
norm of $\Theta - E\Theta$ ⟺ σ_Θ
norm of $X - EX$ ⟺ σ_X
$\cos\alpha$ ⟺ $\rho_{\Theta,X}$.

Find a vector $\hat{\Theta} - E\Theta = a(X - EX)$ that minimizes $\|\Theta - \hat{\Theta}\|$

- Clearly $\Theta - \hat{\Theta} \perp X - EX$ minimizes $\|\Theta - \hat{\Theta}\|$, i.e.,

$E((\Theta - \hat{\Theta})(X - EX)) = 0 \Rightarrow E[(\Theta - E\Theta)(X - EX)] = E[(\hat{\Theta} - E\Theta)(X - EX)]$

$\Rightarrow Cov(\Theta, X) = a \operatorname{Var} X \Rightarrow a = \frac{Cov(\Theta, X)}{\operatorname{Var} X}$.

This arguement is called the orthogonality principle.