Module – 6.1

BAYESIAN ESTIMATION

S. Lakshmivarahan

School of Computer Science

University of Oklahoma

Norman, Ok – 73069, USA

varahan@ou.edu

BAYESIAN FRAMEWORK

- $x \in R^n$ – unknown, random – $p(x)$ is prior
- $z \in R^m$ – observation about x – $p(z|x)$ – conditional distribution
- $\hat{x} = \Phi(z)$ be an estimate
  - Define error $\tilde{x} = x - \hat{x}$
- Cost function $c: R^n \rightarrow R$, $c(\tilde{x})$ is called the cost associated with error
- Properties:
  - $c(0) = 0$
  - $c(a) \leq c(b)$ if $||a|| < ||b||$

EXAMPLES OF COST FUNCTION C(·)

- Sum of squared error

$$c(\tilde{\mathbf{x}}) \;=\; (\mathbf{x}-\hat{\mathbf{x}})^T\,(\mathbf{x}-\hat{\mathbf{x}})$$

- Weighted sum of squared error

$$\begin{aligned} c(\tilde{\mathbf{x}}) &= \tilde{\mathbf{x}}^T\mathbf{W}\tilde{\mathbf{x}} = (\mathbf{x}-\hat{\mathbf{x}})^T\mathbf{W}(\mathbf{x}-\hat{\mathbf{x}}) \\ &= \|(\mathbf{x}-\hat{\mathbf{x}})\|^2_{\mathbf{W}} \end{aligned}$$

- Uniform cost

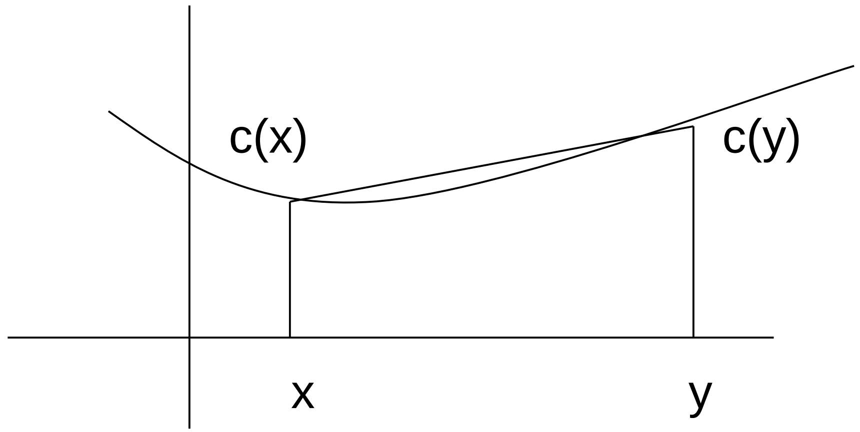
$$c(\tilde{\mathbf{x}}) = \begin{cases} 0, & \text{if } \|\mathbf{x}\| \le \epsilon \\ 1, & \text{otherwise} \end{cases}$$

- Absolute error (x is a scalar)

$$c(\tilde{x}) \;=\; |(x-\hat{x})|$$

EXAMPLES OF COST FUNCTION C(.) (CONT'D)

- Symmetry + Convexity
  - Symmetry: $c(\tilde{\mathbf{x}}) \;=\; c(-\tilde{\mathbf{x}})$
  - Convex: $c(a\mathbf{x} + (1-a)\mathbf{y}) \;\leq\; a\,c(\mathbf{x}) + (1-a)c(\mathbf{y})$
    - $0 \leq a \leq 1$



STATEMENT OF THE PROBLEM

- Given p(x), p(z|x), z and c(.), goal is to minimize Bayes cost function:

$$B(\hat{\mathbf{x}}) = E[c(\tilde{\mathbf{x}})] = \int_{\mathcal{R}^m} \int_{\mathcal{R}^n} c(\mathbf{x} - \hat{\mathbf{x}}) \, \boxed{p(\mathbf{x}, \mathbf{z})} d\mathbf{x} d\mathbf{z}$$

Joint distribution

- Since p(x,z) = p(z|x)p(x) = p(x|z)p(z),

$$B(\hat{\mathbf{x}}) = \int_{\mathcal{R}^m} B(\hat{\mathbf{x}}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z}$$

where

$$B(\hat{\mathbf{x}}|\mathbf{z}) = \int_{\mathcal{R}^n} c(\mathbf{x} - \hat{\mathbf{x}}) p(\mathbf{x}|\mathbf{z}) d\mathbf{x}$$

- Since p(z) > 0, minimizing B($\hat{x}|z$), minimizes B($\hat{x}$)

SPECIAL CASES
A) BAYES LEAST SQUARES ESTIMATOR

- Define $\mu = E[\mathbf{x}|\mathbf{z}] = \int_{\mathcal{R}^n} \mathbf{x}\ p(\mathbf{x}|\mathbf{z}) d\mathbf{x}$
- μ is a function of the observations z
- Then choosing c(x) = (x - $\hat{x}$)^{T}W(x - $\hat{x}$)

$$
\begin{aligned}
B(\hat{\mathbf{x}}) &= E[c(\tilde{\mathbf{x}})] \\
&= E[(\mathbf{x}-\hat{\mathbf{x}})^T \mathbf{W} (\mathbf{x}-\hat{\mathbf{x}})] \\
&= E[(\mathbf{x}-\mu+\mu-\hat{\mathbf{x}})^T \mathbf{W} (\mathbf{x}-\mu+\mu-\hat{\mathbf{x}})] \\
&= E[(\mathbf{x}-\mu)^T \mathbf{W} (\mathbf{x}-\mu)] + E[(\mu-\hat{\mathbf{x}})^T \mathbf{W} (\mu-\hat{\mathbf{x}})] \\
&\quad +2E[(\mathbf{x}-\mu)^T \mathbf{W} (\mu-\hat{\mathbf{x}})]
\end{aligned}
$$

- Using iterated law of Conditional expectations

$$E[(\mathbf{x}-\mu)^T\mathbf{W}(\mu-\hat{\mathbf{x}})] \quad = \quad E\left\{E[(\mathbf{x}-\mu)^T\mathbf{W}(\mu-\hat{\mathbf{x}})|\mathbf{z}]\right\}$$

- But

$$\begin{aligned} E[(\mathbf{x}-\mu)^T\mathbf{W}(\mu-\hat{\mathbf{x}})|\mathbf{z}] &= (\mu-\hat{\mathbf{x}})^T\mathbf{W}E[(\mathbf{x}-\mu)|\mathbf{z}] \\ &= (\mu-\hat{\mathbf{x}})^T\mathbf{W}\{E(\mathbf{x}|\mathbf{z})-\mu\} \\ &= 0 \end{aligned}$$

- $B(\hat{\mathbf{x}}) \quad = \quad E[(\mathbf{x}-\mu)^T \mathbf{W}(\mathbf{x}-\mu)] + E[(\mu-\hat{\mathbf{x}})^T \mathbf{W}(\mu-\hat{\mathbf{x}})]$
- The only control we have is the choice of $\hat{x}$
- B($\hat{x}$) is minimum when

  $\hat{x}$ = μ = E(x|z) = posterior mean

$$
\begin{aligned}
\hat{\mathbf{x}}_{MS} &= E[\mathbf{x}|\mathbf{z}] \\
&= \int_{\mathcal{R}^n} \mathbf{x}\left(\frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{z})}\right)d\mathbf{x} \\
&= \frac{\int_{\mathcal{R}^n} \mathbf{x}p(\mathbf{z}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{R}^n} p(\mathbf{z}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}
\end{aligned}
$$

PROPERTIES OF BAYES LEAST SQUARES ESTIMATE

- $\hat{x}_{MS}$ is unbiased: $$\begin{aligned} E[\mathbf{x} - \hat{\mathbf{x}}_{MS}] &= E\{E[\mathbf{x} - \hat{\mathbf{x}}_{MS}|\mathbf{z}]\} \\ &= E\{E[\mathbf{x}|\mathbf{z}] - \hat{\mathbf{x}}_{MS}\} \\ &= 0 \end{aligned}$$
- $\tilde{x}$ = x - $\hat{x}_{MS}$ => E($\tilde{x}$) = E[x- $\hat{x}_{MS}$] = 0
  - * Mean of the error is zero
- $$B(\hat{\mathbf{x}}_{MS}|\mathbf{z}) = \int_{\mathcal{R}^n} (\hat{\mathbf{x}} - \hat{\mathbf{x}}_{MS})^T (\hat{\mathbf{x}} - \hat{\mathbf{x}}_{MS}) p(\mathbf{x}|\mathbf{z}) d\mathbf{x}$$ [W = I]

  = total variance in the components of $\tilde{x}$
  - Since $\hat{x}_{MS}$ minimizes B($\hat{x}$|z) => $\hat{x}_{MS}$ also minimizes the variance in the estimate

See the similarity

Using only observation

$\hat{x}_{LS}$ is BLUE

(GAUSS-MARKOV)

obs + prior

$\hat{x}_{MS}$ is BLUE

(Bayes Posterior Mean Square)

EXAMPLE 16.2.1

- $z = x + v \quad v \sim N(0, \sigma_v^2) \quad x \sim N(m_x, \sigma_x^2)$, x and v incorrelated

  $E(xv) = 0$

  $\Rightarrow z \sim N(m_x, \sigma^2)$ where $\sigma^2 = \sigma_x^2 + \sigma_v^2$
- Compute $p(x|z)$
  - Recall $p(z|x) = N(x, \sigma_v^2)$, $p(x) = N(m_x, \sigma_x^2)$

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{z}) &= \frac{p(z|x)p(x)}{p(z)} \\
&= \frac{N(x,\sigma_v^2)\;N(\mathbf{m}_x,\sigma_x^2)}{N(\mathbf{m}_x,\sigma^2)} \\
&= \beta\;\exp\{-\frac{1}{2}[\frac{(z-x)^2}{\sigma_v^2} + \frac{(x-\mathbf{m}_x)^2}{\sigma_x^2} - \frac{(z-\mathbf{m}_x)^2}{\sigma^2}]\}
\end{aligned}
$$

EXAMPLE CONTINUED

- Simplifying the term in square brackets:

$$\frac{(z-x)^2}{\sigma_v^2} + \frac{(x-\mathbf{m}_x)^2}{\sigma_x^2} - \frac{(z-\mathbf{m}_x)^2}{\sigma^2} \quad = \quad x^2[\frac{1}{\sigma_v^2} + \frac{1}{\sigma_x^2}] - 2x[\frac{z}{\sigma_v^2} + \frac{\mathbf{m}_x}{\sigma_x^2}]$$

$$+[\frac{z^2}{\sigma_v^2} + \frac{\mathbf{m}_x^2}{\sigma_x^2} - \frac{(z-\mathbf{m}_x)^2}{\sigma^2}] \qquad \text{-> (1)}$$

- We need to compute it as **a perfect square**
- Define

$$\frac{1}{\sigma_e^2} \quad = \quad \frac{1}{\sigma_v^2} + \frac{1}{\sigma_x^2} \quad = \quad \frac{\sigma_v^2+\sigma_x^2}{\sigma_v^2\sigma_x^2} \qquad \text{-> (2)}$$

and

$$\frac{\hat{\mathbf{x}}_{MS}}{\sigma_e^2} = \frac{z}{\sigma_v^2} + \frac{\mathbf{m}_x}{\sigma_x^2} \qquad \text{-> (3)}$$

EXAMPLE CONTINUED

- R.H.S. (1) becomes

$$\frac{1}{\sigma_e^2}[x^2 - 2x\hat{\mathrm{x}}_{MS} + \hat{\mathrm{x}}_{MS}^2] \quad = \quad \frac{1}{\sigma_e^2}(x - \hat{\mathrm{x}}_{MS})^2 \qquad \text{-> (4)}$$

- $\therefore \; p(\mathbf{x}|\mathbf{z}) \quad = \quad \alpha \;\exp[-\frac{1}{2}\frac{(x-\hat{\mathrm{x}}_{MS})^2}{\sigma_e^2}]$
- $\therefore$ Posterior mean is

$$\begin{aligned}
\hat{\mathrm{x}}_{MS} \quad &= \quad (\frac{\sigma_e^2}{\sigma_x^2})\mathbf{m}_x + (\frac{\sigma_e^2}{\sigma_v^2})z \qquad \text{-> (5)} \\
&= \quad (\frac{\sigma_v^2}{\sigma_x^2+\sigma_v^2})\mathbf{m}_x + (\frac{\sigma_x^2}{\sigma_x^2+\sigma_v^2})z \\
&= \quad \alpha\mathbf{m}_x + (1-\alpha)z \qquad \text{-> (6)}
\end{aligned}$$

EXAMPLE CONTINUED

- $\hat{x}_{MS} = m_x + \left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_v^2}\right)[\, z - m_x]$ Kalman-filter form

  prior Kalman gain innovations

- $\therefore$ If $\sigma_x^2 >> \sigma_v^2$ => observation has a larger weight

  If $\sigma_x^2 << \sigma_v^2$ => prior has a larger weight

- $\hat{x}_{MS} = am_x + (1\text{-}a)z$

  $= m_x + am_x - m_x + (1\text{-}a)z$

  $= m_x + (1\text{-}a)[z\text{-}m_x]$ -> (7)

Adaptive nature!

VECTOR CASE: Z = HX + V

- v: $E(v) = 0 \quad E(vv^T) = \Sigma_v \quad v \sim N(0, \Sigma_v)$
- x: $E(x) = m_x \quad cov(x) = \Sigma_x \quad x \sim N(m_x, \Sigma_x)$
  - $\Rightarrow Hx \sim N(Hm_x, H\Sigma_x H^T)$
  - Note: v, x are uncorrelated
- z is normal
- $E(z) = Hm_x$ -> (8)
- $cov(z) = E[(z-Hm_x)(z-Hm_x)^T]$

$$= E[(Hx+v-Hm_x)(Hx+v-Hm_x)^T]$$

$$= E[H(x-m_x)+v][H(x-m_x)+v]^T$$

$$= H[E[(x-m_x)(x-m_x)^T]H^T + E[vv^T]$$

$$= H\,\Sigma_x H^{T} + \Sigma_v \text{ -> (9)}$$

VECTOR CASE

- $E[z|x] = E[\ Hx+v\ |x]$

  $= Hx + E[v|x]$

  $= Hx$ -> (10)
- $cov(z|x) = E(vv^T) = \Sigma_v$ -> (11)
- $p(\ z|x) = N(Hx, \Sigma_v)$ -> (12)

POSTERIOR ANALYSIS

- $p(x|z) = p(z|x)p(x)/p(z)$

  $= N(Hx, \Sigma_v)\, N(m_x, \Sigma_x) / N(Hm_x, \Sigma)$

  $= \alpha \exp[-\tfrac{1}{2}(z-Hx)^T \Sigma_v^{-1}(z-Hx)$

  $- \tfrac{1}{2}(x-m_x)^T \Sigma_x^{-1}(x-m_x)$

  $+ \tfrac{1}{2}(z-Hm_x)^T \Sigma^{-1}(z-Hm_x)$ -> (13)

- Consider the exponent:
  - $x^T[H^T\Sigma_v^{-1}H+\Sigma_x^{-1}]x - 2[H^T\Sigma_v^{-1}z+\Sigma_x^{-1}m_x]^T x + z^T\Sigma_v^{-1}z +$

    $m_x^T\Sigma_x^{-1}m_x - (z-Hm_x)^T\Sigma^{-1}(z-Hm_x)$ -> (14)

    $\equiv (x - \hat{x}_{MS})^T \Sigma_e^{-1} (x - \hat{x}_{MS})$ -> (15)
  - => $\Sigma_e^{-1} = (H^T\Sigma_v^{-1}H+\Sigma_x^{-1})$ -> (16)
  - $\hat{x}_{MS} = \Sigma_e[H^T\Sigma_v^{-1}z+\Sigma_x^{-1}m_x]$ -> (17)

EXERCISES

- Substituting (16), (17) in (15), verify that (14) and (15) are equivalent

REFERENCES

- J. L. Melsa and D. L. Cohn (1978) Decision and Estimation Theory, *McGraw Hill*
- Also refer to chapter 16 in LLD (2006)