ECE531 Screencast 5.2: MMSE Bayesian Estimation

D. Richard Brown III

Worcester Polytechnic Institute

Minimum Mean Squared Error with Scalar Parameter

Squared error cost assignment: $C_\Theta(g(y)) = (g(y) - \Theta)^2$.

We want to minimize the posterior cost

$$\hat{\theta}_{\mathsf{opt}}(y) = \arg\min_{g(\cdot)} \mathrm{E}[C_\Theta(g(y)) \,|\, Y = y] = \arg\min_{g(\cdot)} \mathrm{E}[(g(y) - \Theta)^2 \,|\, Y = y]$$

Note that y is fixed. Hence $g(y) = u$ is also fixed and

$$\begin{aligned} \hat{\theta}_{\mathsf{mmse}}(y) &= \arg\min_{g(\cdot)} \mathrm{E}[(g(y) - \Theta)^2 \,|\, Y = y] \\ &= \arg\min_{u} u^2 - 2u\mathrm{E}[\Theta \,|\, Y = y] + \mathrm{E}[\Theta^2 \,|\, Y = y] \end{aligned}$$

We can find the minimum by taking a derivative with respect to u and setting it equal to zero...

$$\frac{\partial}{\partial u}\left\{u^2 - 2u\mathrm{E}[\Theta \,|\, Y = y] + \mathrm{E}[\Theta^2 \,|\, Y = y]\right\} = 2u - 2\mathrm{E}[\Theta \,|\, Y = y] = 0$$

hence $\hat{\theta}_{\mathsf{mmse}}(y) = \mathrm{E}[\Theta \,|\, Y = y]$. The MMSE estimator is just the conditional mean of the random parameter Θ given the observation $Y = y$.

Example: Estimation of a Constant in White Noise

Suppose we observe

$$Y_k = \Theta + W_k \qquad k = 0, \dots, n-1$$

where $W \sim \mathcal{N}(0, \sigma^2 I)$ and $\Theta \sim \mathcal{N}(\mu, v^2)$. Note that Θ is a scalar parameter.

Note that v^2 is a measure of the accuracy of our prior knowledge. If v^2 is small, we know Θ accurately without any observations.

Solution steps:

1. Use the observation model to determine the conditional distribution $p_\theta(y)$.
2. Use Bayes' rule to determine the posterior distribution $\pi_y(\theta)$.
3. Compute the conditional mean $\hat{\theta}_{\mathsf{mmse}}(y) = \mathrm{E}[\Theta \,|\, Y = y]$.

See Example 10.1 in your textbook for the details...

Example: Estimation of a Constant in White Noise

$$\hat{\theta}_{\text{mmse}}(y) = \mathrm{E}[\Theta \,|\, Y = y] = \frac{\frac{v^2}{\sigma^2} n\bar{y} + \mu}{\frac{v^2}{\sigma^2} n + 1}$$

$$\text{MMSE} = \mathrm{E}\left[\text{var}[\Theta \,|\, Y = y]\right] = \frac{v^2}{\frac{v^2}{\sigma^2} n + 1}$$

where $\bar{y} := \frac{1}{n}\sum_{k=0}^{n-1} y_k$. Remarks:

- When $n = 0$, the MMSE estimate $\hat{\theta} = \mu$ and the MMSE is simply v^2.
- Note that MMSE is strictly decreasing in n as long as $v > 0$.
- The effect of the prior on $\hat{\theta}_{\text{mmse}}$ also becomes less important with more samples. In the limit

$$\lim_{n\to\infty} \hat{\theta}_{\text{mmse}} = \bar{y} \qquad \text{and} \qquad \lim_{n\to\infty} \text{MMSE} = 0$$

Minimum Mean Squared Error with Vector Parameter

Squared error cost assignment: $C_\Theta(g(y)) = \|g(y) - \Theta\|_2^2$.

Note that y is fixed. Hence $g(y) = u$ is also fixed and

$$\begin{aligned}\hat{\theta}_{\mathsf{mmse}}(y) &= \arg\min_{g(\cdot)} \mathrm{E}[\|g(y) - \Theta\|_2^2 \,|\, Y = y] \\ &= \arg\min_{u} u^\top u - 2u^\top \mathrm{E}[\Theta \,|\, Y = y] + \mathrm{E}[\Theta^\top \Theta \,|\, Y = y]\end{aligned}$$

How do we solve this sort of problem? We can find the minimum by taking the gradient with respect to u and setting it equal to zero...

$$\nabla_u \left\{ u^\top u - 2u^\top \mathrm{E}[\Theta \,|\, Y = y] + \mathrm{E}[\Theta^\top \Theta \,|\, Y = y] \right\} = 2u - 2\mathrm{E}[\Theta \,|\, Y = y]$$

hence

$$2u = \mathrm{E}[2\Theta \,|\, Y = y] \qquad \Leftrightarrow \qquad u = \mathrm{E}[\Theta \,|\, Y = y]$$

and we can conclude that $\hat{\theta}_{\mathsf{mmse}}(y) = \mathrm{E}[\Theta \,|\, Y = y]$.

Performance of Bayesian MMSE Estimator

$$\text{MMSE} = \mathrm{E}\left[\|\Theta - \hat{\theta}_{\text{mmse}}(Y)\|_2^2\right]$$

where the expectation is evaluated with respect to the joint pdf $p_{Y,\Theta}(y,\theta)$.

$$\begin{aligned}
\text{MMSE} &= \int\int \|\theta - \mathrm{E}[\Theta \,|\, Y = y]\|_2^2 \, p_{Y,\Theta}(y,\theta)\, dy\, d\theta \\
&= \int\int \|\theta - \mathrm{E}[\Theta \,|\, Y = y]\|_2^2 \, \pi_y(\theta)\, d\theta\, p(y)\, dy \\
&= \int\int \sum_i \left(\theta_i - \mathrm{E}[\Theta_i \,|\, Y = y]\right)^2 \, \pi_y(\theta)\, d\theta\, p(y)\, dy \\
&= \int \sum_i \text{var}(\Theta_i \,|\, Y = y)\, p(y)\, dy \\
&= \int \text{trace}\left\{\text{cov}(\Theta \,|\, Y = y)\right\} \, p(y)\, dy
\end{aligned}$$

where trace$(\cdot)$ is the sum of the diagonal elements of a matrix.