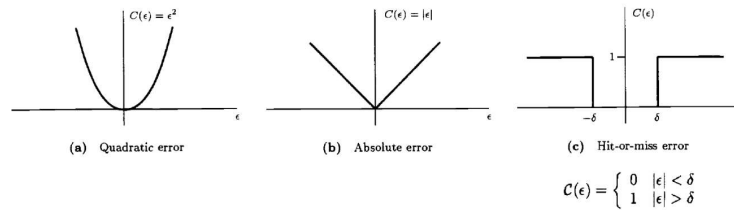


MAP

1. Bayes Risk

error에 대한 cost function에 따라 MMSE가 달라질 수 있다. 여기에서는 대표적으로 세 개의 cost function에 대해서 살펴보자.

Various Cost Function



1) Quadratic error : $C(e) = e^2$

이전에 구했듯이 MMSE는 $\hat{\theta} = E(\theta|x)$ 이다.

2) Absolute error : $C(e) = |e|$

미분을 하여 0이라고 두면

$$\begin{aligned} g(\hat{\theta}) &= \int |\theta - \hat{\theta}| p(\theta|x) d\theta \\ &= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|x) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta|x) d\theta \\ \frac{dg(\hat{\theta})}{d\hat{\theta}} &= \int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|x) d\theta = 0 \\ \int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta &= \int_{\hat{\theta}}^{\infty} p(\theta|x) d\theta \end{aligned}$$

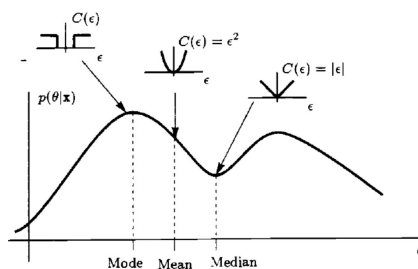
따라서 MMSE는 posterior pdf의 median 값을 가진다.

3) Hit or miss error

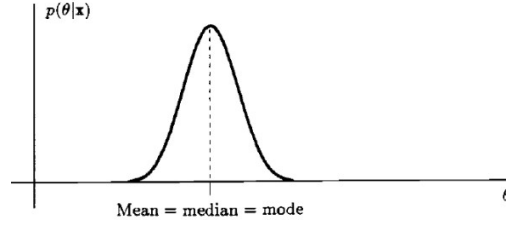
$$\begin{aligned} g(\hat{\theta}) &= \int_{-\infty}^{\hat{\theta}-\delta} 1 \cdot p(\theta|x) d\theta + \int_{\hat{\theta}+\delta}^{\infty} 1 \cdot p(\theta|x) d\theta \\ &= 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|x) d\theta \end{aligned}$$

$g(\hat{\theta})$ 가 최대가 되기 위해선 posterior pdf가 최대가 되어야 한다. 따라서 구하고자 하는 MMSE는 $\hat{\theta} = \arg \max p(\theta|x)$ 가 되며 이를 maximum a posteriori(MAP)라고 한다.

general한 posterior pdf에서 각 cost별로 MMSE는 다음과 같이 나타난다.



그리고 그림이 세 MMSE 값이 일치하는 pdf가 gaussian 분포이다.



2. Bayesian Linear Model

다음과 같은 데이터가 주어졌다고 하자.

$$x[n] = A + w[n]$$

여기서 A 가 선형 모델이라고 생각하면 다음과 같다.

$$x[n] = H\theta + w[n]$$

위 모델은 두 확률변수 x , θ 가 결합분포를 이루므로 그에 따라 posterior pdf를 구할 수 있다. z 를 다음과 같이 정의하자.

$$z = \begin{bmatrix} H\theta + w \\ \theta \end{bmatrix}$$

그러면 각각의 기댓값은 다음과 같다.

$$E(x) = H\mu_\theta$$

$$E(y) = E(\theta) = \mu_\theta$$

또한 공분산은 다음과 같다.

$$\begin{aligned} C_{xx} &= E((x - E(x))(x - E(x))^T) \\ &= E((H\theta + w - H\mu_\theta)(H\theta + w - H\mu_\theta)^T) \\ &= E((H(\theta - \mu_\theta) + w)(H(\theta - \mu_\theta) + w)^T) \\ &= HE((\theta - \mu_\theta)(\theta - \mu_\theta)^T)H^T + E(ww^T) \\ &= HC_\theta H^T + C_w \end{aligned}$$

$$\begin{aligned} C_{yx} &= E((y - E(y))(x - E(x))^T) \\ &= E((\theta - \mu_\theta)(H(\theta - \mu_\theta) + w)^T) \\ &= E((\theta - \mu_\theta)(H(\theta - \mu_\theta))^T) \\ &= C_\theta H^T \end{aligned}$$

따라서 이를 결합분포의 기댓값과 분산 식에 대입하면 다음과 같다.

$$E(\theta|x) = \mu_\theta + C_\theta H^T (HC_\theta H^T + C_w)^{-1} (x - H\mu_\theta)$$

$$C_{\theta|x} = C_\theta - C_\theta H^T (HC_\theta H^T + C_w)^{-1} HC_\theta$$

위 식을 앞서 소개했던 다음의 Matrix inversion lemma를 사용하여 변환할 수 있다.

$$\begin{aligned} (A_{11} + A_{12}A_{22}^{-1}A_{21})^{-1} &= A_{11}^{-1} - A_{11}^{-1}A_{12}(A_{22} + A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} \\ (A_{11} + A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} &= A_{11}^{-1}A_{12}(A_{22} + A_{21}A_{11}^{-1}A_{12})^{-1} \end{aligned}$$

$$E(\theta|x) = \mu_\theta + (C_\theta^{-1} + H^T C_w^{-1} H)^{-1} H^T C_w^{-1} (x - H\mu_\theta)$$

$$C_{\theta|x} = (C_\theta^{-1} + H^T C_w^{-1} H)^{-1}$$

만약 위 식에서 사전 지식이 주어지지 않는다면, $\mu_\theta \rightarrow 0, C_\theta^{-1} \rightarrow 0$ 이 되어 선형 모델의 MVUE와 같아진다.

$$\hat{\theta} = E(\theta|x) = (H^T C_w^{-1} H)^{-1} H^T C_w^{-1} x$$

3. MAP

MAP는 다음과 같이 구할 수 있다.

$$\begin{aligned}\hat{\theta} &= \arg \max p(\theta|x) \\ &= \arg \max p(x|\theta)p(\theta) \\ &= \arg \max [\log p(x|\theta) + \log p(\theta)]\end{aligned}$$

Ex) Exponential PDF

다음과 같은 pdf가 주어졌을 때,

$$p(x[n]|\theta) = \begin{cases} \theta \exp(-\theta x[n]) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$p(\theta) = \begin{cases} \lambda \exp(-\lambda \theta) & \theta > 0 \\ 0 & \theta < 0 \end{cases}$$

$p(x|\theta)$ 는 다음과 같다.

$$p(x|\theta) = \theta^N \exp(-\theta N\bar{x})$$

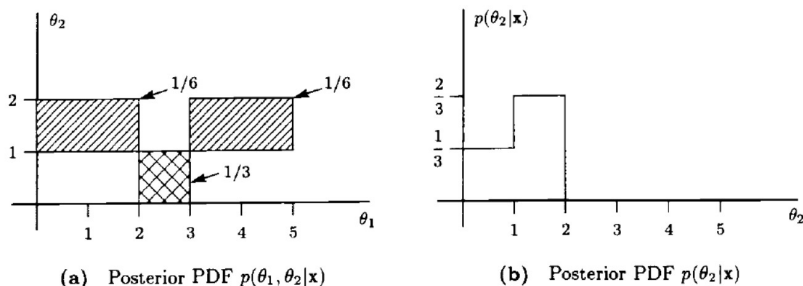
따라서 다음의 $g(\theta)$ 의 최대일 때를 구해야 한다. 따라서 이를 미분해 0으로 두어 MAP를 구할 수 있다.

$$\begin{aligned}g(\theta) &= \ln p(x|\theta) + \ln p(\theta) \\ &= \ln[\theta^N \exp(-\theta N\bar{x})] + \ln[\lambda \exp(-\lambda \theta)] \\ &= N \ln \theta - N\theta \bar{x} + \ln \lambda - \lambda \theta\end{aligned}$$

$$\frac{dg(\theta)}{d\theta} = \frac{N}{\theta} - N\bar{x} - \lambda = 0$$

$$\hat{\theta} = \frac{1}{\bar{x} + \frac{\lambda}{N}}$$

4. Scalar MAP vs Vector MAP



(a)를 θ_2 를 기준으로 그린 것이 (b)이다.

(a)에서의 MAP estimator는 $2 < \hat{\theta}_1 < 3, 0 < \hat{\theta}_2 < 1$ 임을 알 수 있다.

반면, (b)에서는 MAP estimator가 $1 < \hat{\theta}_2 < 2$ 임을 알 수 있다.

이처럼 같은 posterior에 대해서도 스칼라와 벡터가 다른 MAP를 가진다.

5. Invariance Property of the MLE

이전에 MLE에 대해서 다음과 같은 정리가 성립한다는 것을 알았다.

Theorem 7.2 (Invariance Property of the MLE) The MLE of the parameter $\alpha = g(\theta)$, where the PDF $p(\mathbf{x}; \theta)$ is parameterized by θ , is given by

$$\hat{\alpha} = g(\hat{\theta})$$

where $\hat{\theta}$ is the MLE of θ . The MLE of $\hat{\theta}$ is obtained by maximizing $p(\mathbf{x}; \theta)$. If g is not a one-to-one function, then $\hat{\alpha}$ maximizes the modified likelihood function $\bar{p}_T(\mathbf{x}; \alpha)$, defined as

$$\bar{p}_T(\mathbf{x}; \alpha) = \max_{\{\theta: \alpha = g(\theta)\}} p(\mathbf{x}; \theta).$$

하지만 MAP에 대해서는 성립하지 않는다. 예시를 살펴보자.

이전에 살펴보았던 Exponential PDF 예제를 살펴보자. 이 때의 MAP는 다음과 같았다.

$$\hat{\theta} = \frac{1}{\bar{x} + \frac{\lambda}{N}}$$

그렇다면 과연 $\alpha = \frac{1}{\theta}$ 에 대해서도 $\hat{\alpha} = \bar{x} + \frac{\lambda}{N}$ 이 성립할까? 우선 α 에 대해서 식을 다시 써 보면 다음과 같다.

$$p(x[n]|\alpha) = \begin{cases} \frac{1}{\alpha} \exp(-\frac{x[n]}{\alpha}) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

$$p_\alpha(\alpha) = \frac{p_\theta(\theta(\alpha))}{\left| \frac{d\alpha}{d\theta} \right|} = \begin{cases} \frac{\lambda \exp(-\lambda/\alpha)}{\alpha^2} & \alpha > 0 \\ 0 & \alpha < 0 \end{cases}$$

이에 대해 MAP를 구하면 다음과 같다.

$$\begin{aligned} g(\alpha) &= \ln\left[\left(\frac{1}{\alpha}\right)^N \exp\left(-\frac{1}{\alpha} N\bar{x}\right)\right] + \ln \frac{\lambda \exp(-\lambda/\alpha)}{\alpha^2} \\ &= -N \ln \alpha - N \frac{\bar{x}}{\alpha} + \ln \lambda - \frac{\lambda}{\alpha} - 2 \ln \alpha \\ &= -(N+2) \ln \alpha - \frac{N\bar{x} + \lambda}{\alpha} + \ln \lambda \\ \frac{dg}{d\alpha} &= -\frac{N+2}{\alpha} + \frac{N\bar{x} + \lambda}{\alpha^2} \\ \hat{\alpha} &= \frac{N\bar{x} + \lambda}{N+2} \end{aligned}$$