Problem 3: Maximum Posterior Estimation

- We are given that the target variable are drawn from gaussian distribution,

$$P(y|\underline{x}, w, \beta) = N(y|f(\underline{x}, w), \beta^{-1}) \qquad (1)$$

where input values $x = (x_1, \dots, x_n)^T$, and corresponding target values, $y = (y_1, \dots, y_n)^T$, w is the parameter vector governing the mean and β is the precision.

- Also, the prior gaussian distribution for w,

$$P(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{(M+1/2)} \exp\left(-\frac{\alpha}{2} w^T w\right) \qquad (2)$$

Using Bayes theorem, the posterior probability for ω is proportional to the product of the prior distribution and the likelihood function,

$$p(\omega|x, y, \alpha, \beta) \propto p(y|x, \omega, \beta) \cdot p(\omega|\alpha) \quad \text{--- (3)}$$

Maximizing the posterior distribution, taking logarithm,

$$\log[p(\omega|x, y, \alpha, \beta)] = \log[p(y|x, \omega, \beta) \cdot p(\omega|\alpha)]$$

$$= \log[p(y|x, \omega, \beta)] + \log(p(\omega|\alpha)) \quad \text{--- (4)}$$

The RHS of the equation becomes,

$$RHS = \log\left[\prod_{n=1}^{N} N\left(y|f(x,\omega), \beta^{-1}\right)\right] + \log\left[p(\omega|\alpha)\right].$$

Here, the precision $\beta = 1/\text{Variance} = 1/\sigma^2$, using this, the RHS of the equation becomes,

$$RHS = \log\left[\prod_{n=1}^{N} N\left(y|f(x,\omega), \sigma^{2}\right)\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \log\left[\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{y-f(x,\omega)}{\sigma}\right)^2\right)\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \left[\sum_{n=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2}\left(\frac{y-f(x,\omega)}{\sigma}\right)^2\right] + \log\left[p(\omega/\alpha)\right]$$

$$= \left[\sum_{n=1}^{N}\left(\cancel{\log(1)} - \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\left(y-f(x,\omega)\right)^2\right)\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \left[-\sum_{n=1}^{N} \log\left(\sqrt{2\pi\sigma^2}\right) - \sum_{n=1}^{N} \frac{1}{2\sigma^2}\left(y-f(x,\omega)\right)^2\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \left[\frac{-1}{2\sigma^2}\sum_{n=1}^{N}\left(y-f(x,\omega)\right)^2 - \sum_{n=1}^{N} \log\left(\left(2\pi\sigma^2\right)^{1/2}\right)\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \left[\frac{-1}{2\sigma^2}\sum_{n=1}^{N}\left(y-f(x,\omega)\right)^2 - N\log\left(2\pi\sigma^2\right)^{1/2}\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \left[\frac{-1}{2\sigma^2}\sum_{n=1}^{N}\left(y-f(x,\omega)\right)^2 - \frac{N}{2}\log\left(2\pi\sigma^2\right)\right] + \log\left[p(\omega|\alpha)\right]$$

$$= \left[\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (y - f(x,\omega))^2 - \frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) \right] + \log[p(\omega|\alpha)]$$

$$= \left[-\frac{\beta}{2} \sum_{n=1}^{N} (y - f(x,\omega))^2 - \frac{N}{2} \log(2\pi) + \frac{N}{2} \log(\beta) \right] + \left[\frac{(M+1)}{2} \log\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2} \omega^T \omega \right]$$

So, the equation (4) becomes,

$$\log(p(\omega|x,y,\alpha,\beta)) = \left[-\frac{\beta}{2} \sum_{n=1}^{N} (y - f(x,\omega))^2 - \frac{N}{2} \log(2\pi) + \frac{N}{2} \log(\beta) \right]$$
$$+$$
$$\left[\frac{(M+1)}{2} \log\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2} \omega^T \omega \right]$$

Taking partial derivative w.r.t. ω,

$$\frac{\partial}{\partial \omega} \left(\log(p(\omega|x,y,\alpha,\beta)) \right) = -\frac{\beta}{2} \sum_{n=1}^{N} (y - f(x,\omega))^2 - \frac{\alpha}{2} \omega^T \omega$$

Setting it to 0, we get,

$$0 = -\frac{\beta}{2} \sum_{n=1}^{N} (y - f(x,\omega))^2 - \frac{\alpha}{2} \omega^T \omega$$

Multiply by -1,

$$\frac{\beta}{2} \sum_{n=1}^{N} (y - f(x,\omega))^2 + \frac{\alpha}{2} \omega^T \omega = 0$$

Therefore, $\frac{\beta}{2} \sum_{n=1}^{N} \{ f(x,\omega) - y \}^2 + \frac{\alpha}{2} \omega^T \omega = 0$

Thus, we see maximizing the posterior distribution is equivalent to minimizing the regularized sum-of-squares error function.