Relation between Likelihood function and least squares error under Gaussian noise model

Notation for Gaussian distribution

$$P(\mathbf{t} \mid X, w, \beta^{-1}) = \mathcal{N}(\mathbf{t} \mid y(X, w), \beta^{-1})$$

Targets   Inputs

Target posterior (also known as likelihood function)

$$= \prod_{n=1}^{N} \mathcal{N}(t_n \mid y(x_n, w), \beta^{-1}) \quad \ldots (1)$$

(Due to data independence assumption)

Gaussian distribution: $\mathcal{N}(t_n \mid y(x_n, w), \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}}\, e^{-\frac{\beta}{2}(t_n - y(x_n, w))^2}$

taking log on both sides $\Rightarrow \ln \mathcal{N}(t_n \mid y(x_n, w), \beta^{-1}) = \frac{1}{2}(\ln \beta - \ln(2\pi)) - \frac{\beta}{2}(t_n - y(x_n, w))^2 \quad \ldots (2)$

Substitute result (2) in (1):

$$\ln(P(\mathbf{t} \mid X, w, \beta^{-1})) = \sum_{n=1}^{N} \left(\frac{1}{2}(\ln \beta - \ln(2\pi)) - \frac{\beta}{2}(t_n - y(x_n, w))^2 \right)$$

Log of posterior (Log Likelihood function)

$$= \underbrace{\frac{N}{2}(\ln \beta - \ln(2\pi))}_{\text{Constant wrt } w} - \frac{\beta}{2} \underbrace{\sum_{n=1}^{N} (t_n - y(x_n, w))^2}_{\text{Sum of squares error}}$$

Result: Maximizing log likelihood function is equivalent to minimizing sum of squares error under Gaussian noise model