Maximum Likelihood And Least Squares Method

We take a target value t to be as a deterministic function $y(x, w)$ with additive gaussian noise so that :-

$$t = y(x, w) + \epsilon$$

where ϵ is a zero mean Gaussian random variable with precision (inverse variance) β thus we can write

$$p(t|x, w, \beta) = \mathcal{N}(t|y(x, w), \beta^{-1})$$

Recall that if we assume a squared loss function for a new value of x, will be conditional mean of target value.

In the case of Gaussian distribution of the form, the conditional mean will be :-

$$\mathbb{E}[t|x] = \int t\,p(t|x)\,dt = y(x, w)$$

Now, consider a dataset of inputs $X = \{x_1, x_2 \ldots x_N\}$ with corresponding targets values $\{t_1, t_2 \ldots t_N\}$ We group the target $\{t_n\}$ into a column vector that we denote by t where the typeface is chosen to distinguish it from a single observation of a multivariate target which would be denoted by t. Making the assumption that these data points are drawn independently from the distribution, we obtain the following expressions for the likelihood function which is a function of the adjustable parameters w and β, is the form

$$p(t|x, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n | w^T \phi(x_n), \beta^{-1})$$

Taking the logarithm of the likelihood function, and making use of standard form for the univariate Gaussian, we have

$$\ln [p(t|w, \beta)] = \sum_{n=1}^{N} \ln \mathcal{N}(t_n | w^T \phi(x_n), \beta^{-1})$$

$$\frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \beta E_D(w)$$

Where the sum-of-squares error function is defined by

$$E_D(w) = \frac{1}{2} \sum_{n=1}^{N} \{t_n - w^T \phi(x_n)\}^2$$

We use maximum likelihood to determine w and β. Consider first the maximization with respect to w. As observed already, we see maximization of likelihood function under conditional Gaussian noise distribution for a linear model is equivalent to minimizing sum of squares

$$\nabla \ln p(t|w, \beta) = \sum_{n=1}^{N} \{ t_n - w^T \phi(x_n) \} \phi(x_n)^T$$

setting the gradient to 0 gives us

$$\sum_{n=1}^{N} t_n \phi(x_n)^T - w^T \left(\sum_{n=1}^{N} \phi(x_n) \phi(x_n)^T \right)$$

solving for w we obtain:-

$$\boxed{w_{ML} = (\phi^T \phi)^{-1} \phi^T t}$$