$y = \omega_0 + \omega_1 x_1 + \cdots \omega_n x_n$

These are independent assumption

This (imp) also assumed

$(x_1, t_1)\ (x_2, t_2) \cdots (x_N, t_N)$

$\therefore t \sim N(y(x,w), \beta^{-1})$

$\beta^{-1} = \sigma^{-2}$

$$p(t_1, t_2, \cdots, t_N \mid x_1, x_2 \cdots x_N, \omega)$$

$p(t \mid x, w)$ is likelihood

$$= \prod_{n=1}^{N} p(t_n \mid x_n, w_n) = \prod_{n=1}^{N} \frac{\sqrt{\beta}}{2\pi} e^{-\frac{\beta}{2}(y(x_n,w)-t_n)^2}$$

$$= \frac{\beta^{N/2}}{(2\pi)^{N/2}} e^{-\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n,w)-t_n)^2}$$

$$-\log(p(t_1, \cdots t_N \mid x_1 \cdots x_n, w))$$

$$= -\frac{N}{2}\log\beta + \frac{N}{2}\log 2\pi + \frac{\beta}{2}\sum_{n=1}^{N}(y(x_n,w)-t_n)^2$$

$\min \frac{1}{2}\sum_{n=1}^{N}(y(x_N,w)-t_n)^2$, and you optimal $\hat{y}$

$$\frac{\partial \text{of} \log\cdots}{\partial\beta} = \frac{N}{2}\frac{1}{\beta} + \frac{1}{2}\sum_{n=1}^{N}(y(x_n,w)-t_n)^2 = 0$$

$$= \frac{N}{\beta} = \sum_{n=1}^{N} \cdots$$

$$= \beta = \frac{N}{\sum_{n=1}^{N}(y(x_n,w)-t_n)^2}$$

you get variance

why are we getting a point estimate?

We can get a distribution of $\omega_1, \omega_2 \ldots \omega_n$ and choosing its mean as probability distribution

probability Distribution of coefficients

$$p(\omega_0, \omega_1 \ldots \omega_m \mid x, t)$$

$$p(\underline{\omega} / x, t) \propto p(t / x, \omega)\, p(\omega)$$

$$\because p(t/x) = \int_{\omega} p(t/x, \omega)\, p(\omega)\, d\omega$$

$= \text{constant}$

sum rule in probability

may not be always true, but it is an assumption that works lol. IRL

← Same variance

$$t \sim N(y(x, \omega), \beta^{-1})$$

we already know $p(t / \overset{x,}{\omega})$ [likelihood function]

we dunno the distribution of $p(\omega / x, t)$ (POSTERIOR)

we do however know the distribution of ω without seeing the training examples (PRIOR)

we can call $p(\omega)$ as prior & $p(\omega / x, t)$ posterior