Given that

$$E_D = \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_n)\}^2 \tag{3.104}$$

we find the $\boldsymbol{w}_{\mathrm{ML}}$ by searching the point nulling the partial derivative of E_D w.r.t $\boldsymbol{w}$

$$\frac{\partial E_D}{\partial \boldsymbol{w}} = \sum_{n=1}^{N} r_n \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_n)\boldsymbol{\phi}(\boldsymbol{x}_n)^\top - r_n t_n \boldsymbol{\phi}(\boldsymbol{x}_n)^\top$$

$$\begin{aligned}
& \sum_{n=1}^{N} r_n \boldsymbol{w}_{\mathrm{ML}}^\top \boldsymbol{\phi}(\boldsymbol{x}_n)\boldsymbol{\phi}(\boldsymbol{x}_n)^\top - r_n t_n \boldsymbol{\phi}(\boldsymbol{x}_n)^\top = \mathbf{0} \\
\Leftrightarrow \quad & \boldsymbol{w}_{\mathrm{ML}}^\top \sum_{n=1}^{N} \boldsymbol{\phi}(\boldsymbol{x}_n) r_n \boldsymbol{\phi}(\boldsymbol{x}_n)^\top = \sum_{n=1}^{N} r_n t_n \boldsymbol{\phi}(\boldsymbol{x}_n)^\top \\
\Leftrightarrow \quad & \boldsymbol{w}_{\mathrm{ML}}^\top \boldsymbol{\Phi}^\top \operatorname{diag}(\boldsymbol{r})\boldsymbol{\Phi} = \mathsf{t}^\top \operatorname{diag}(\boldsymbol{r})\boldsymbol{\Phi} \\
\Leftrightarrow \quad & \boldsymbol{\Phi}^\top \operatorname{diag}(\boldsymbol{r})\boldsymbol{\Phi}\boldsymbol{w}_{\mathrm{ML}} = \boldsymbol{\Phi}^\top \operatorname{diag}(\boldsymbol{r})\mathsf{t} \\
\Leftrightarrow \quad & \boldsymbol{w}_{\mathrm{ML}} = (\boldsymbol{\Phi}^\top \operatorname{diag}(\boldsymbol{r})\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \operatorname{diag}(\boldsymbol{r})\mathsf{t}
\end{aligned}$$

From equations (3.11-12), the noise variance of $\boldsymbol{x}_n$ is scaled by r_n from β^{-1} to $(\beta r_n)^{-1}$. Also, one can interpret it as a leverage of data point's importance by attributing a larger weight to frequently appearing data. ■