3. (a) $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$, $R = diag(r_1, \cdots r_n)_{n \times n}$

Then $(Xw - t)^T R (Xw - t)$

$$= (w^T - t^T) R (Xw - t) = (w^T (x_1, \cdots x_n) - (t_1, \cdots t_n)) \begin{pmatrix} \frac{r_1}{2} & & \\ & \ddots & \\ & & \frac{r_n}{2} \end{pmatrix} \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} w - \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \right)$$

$$= ((w^T x_1, \cdots w^T x_n) - (t_1, \cdots t_n)) \begin{pmatrix} \frac{r_1}{2} & & \\ & \ddots & \\ & & \frac{r_n}{2} \end{pmatrix} \left(\begin{pmatrix} x_1 w \\ \vdots \\ x_n w \end{pmatrix} - \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \right)$$

$$= ((w^T x_1 - t_1, \cdots w^T x_n - t_n)) \begin{pmatrix} \frac{r_1}{2} & & \\ & \ddots & \\ & & \frac{r_n}{2} \end{pmatrix} \begin{pmatrix} x_1 w - t_1 \\ \vdots \\ \cancel{x_n w t_n} \\ x_n w - t_n \end{pmatrix}$$

$$= \left(\frac{r_1}{2}(w^T x_1 - t_1) \cdots \frac{r_n}{2}(w^T x_n - t_n) \right) \begin{pmatrix} x_1 w t_1 \\ \vdots \\ x_n w - t_n \end{pmatrix} = \frac{1}{2} \sum r_i (w^T x_i\, t_i)^2$$

(b) $E_D(w) = (Xw - t)^T R (Xw - t)$

From $\frac{\partial E_D(w)}{\partial w} = 2X^T R (Xw - t) = 0 \Rightarrow X^T R X w - X^T R t = 0$

$\Rightarrow w^* = (X^T R X)^{-1} X^T R t$

(c) $\ln P(t_i | x_i; w) = -\frac{1}{2} \ln 2\pi - \ln \sigma_i - \frac{(t_i - w^T x_i)^2}{2\sigma_i^2}$

$$L = \ln \prod_{i=1}^{n} P(t_i | x_i, w) = \sum \ln P(t_i | x_i; w)$$
$$= -\frac{n}{2} \ln 2\pi - \sum_{i=1}^{n} \ln \sigma_i - \sum_{i=1}^{n} \frac{(t_i - w^T x_i)^2}{2(\sigma_i)^2}$$

$$\nabla \ln P(t | x; w) = \frac{\partial L}{\partial w} = \sum \frac{1}{(\sigma_i)^2} (t_i - w^T x_i) x_i = 0 \Rightarrow r_i = \frac{1}{\sigma_i^2}$$

5. (a) $E(w) = -\ln P(t|w)$

$$= -\ln \left[\prod_{n=1}^{N} \prod_{k=0}^{K-1} p(C_k|\phi(x_n))^{\cancel{1\{t_n=k\}}} \cdot 1\{t_n=k\} \right]$$

$$= -\left[\sum_{1}^{N} \sum_{0}^{K-1} 1_{\{t_n=k\}} \log \frac{\exp\{w_k^T \phi(x_n)\}}{\sum_{k=0}^{K-1} \exp\{w_k^T \phi(x_n)\}} \right]$$

Therefore, $\nabla_{w_j} E(w) = \frac{\partial E}{\partial W} = \partial \left[-\sum_{1}^{N} \sum_{0}^{K-1} 1_{\{t_n=k\}} \log \frac{\exp\{w_k^T \phi(x_n)\}}{\sum_{k=0}^{K-1} \exp\{w_k^T \phi(x_n)\}} \right] / \partial W$

$$= \frac{\partial \left[-\sum_{1}^{N} \sum_{0}^{K-1} 1_{\{t_n=k\}} [w_k^T \phi(x_n) - \log \sum_{k=0}^{K-1} \exp(w_k^T \phi(x_n)) \right]}{\partial W}$$

$$= -\sum_{1}^{N} \left[\phi(x_n) \left(1_{\{t_n=j\}} - \log \frac{\exp\{w_j^T \phi(x_n)\}}{\sum_{k=0}^{K-1} \exp\{w_k^T \phi(x_n)\}} \right) \right]$$

(b) $E^{\lambda}(w) = E(w) + \frac{\lambda}{2} \sum_{k=0}^{K-1} w_k^T w_k$

$$\therefore \nabla_{w_j} E^{\lambda}(w) = -\sum_{1}^{N} \left[\phi(x_n) \left(1_{\{t_n=j\}} - \log \frac{\exp\{w^T \phi(x_n)\}}{\sum_{0}^{K-1} \exp} \right) \right] + \lambda w_j$$