109061621 盧冠維 電機碩一 Homework 3

5.6

Known that $y_k = \sigma(a_k)$, where $\sigma(\cdot)$ represent the logistic sigmoid function. Moreover, $\frac{d\sigma}{da} = \sigma(1-\sigma)$

$$\frac{dE(w)}{da_k} = -t_k \frac{1}{y_k}\left[y_k(1-y_k)\right] + (1-t_k)\frac{1}{1-y_k}\left[y_k(1-y_k)\right] = \left[y_k(1-y_k)\right]\left[\frac{1-t_k}{1-y_k} - \frac{t_k}{y_k}\right] = (1-t_k)y_k - t_k(1-y_k) = y_k - t_k$$

5.24

It's obvious. According to (5.113), we have:

$$\tilde{a}_j = \sum_i \tilde{w}_{ji}\tilde{x}_i + \tilde{w}_{j0} = \sum_i \frac{1}{a} w_{ji}\cdot(ax_i + b) + w_{j0} - \frac{b}{a}\sum_i w_{ji}$$
$$= \sum_i w_{ji}x_i + w_{j0} = a_j$$

Where we have used (5.115), (5.116) and (5.117). Currently, we have proved that under the transformation the hidden unit a_j is unchanged. If the activation function at the hidden unit is also unchanged, we have $\tilde{z}_j = z_j$. Now we deal with the output unit $\tilde{y}_k$:

$$\tilde{y}_k = \sum_j \tilde{w}_{kj}\tilde{z}_j + \tilde{w}_{k0} = \sum_j c\tilde{w}_{kj}\tilde{z}_j + c\tilde{w}_{k0} + d$$
$$= c\sum_j\left[\tilde{w}_{kj}\tilde{z}_j + \tilde{w}_{k0}\right] + d = cy_k + d$$

Where we have used (5.114), (5.119) and (5.120). To be more specific, here we have proved that the linear transformation between $\tilde{y}_k$ and y_k can be achieved by making transformation (5.119) and (5.120).

6.16

Based on the total derivative of function f, we have:

$$f\left((w+\Delta w)^T\phi_1, (w+\Delta w)^T\phi_2, \cdots, (w+\Delta w)^T\phi_N\right) = \sum_{n=1}^{N}\frac{\partial f}{\partial(w^T\phi_n)}\cdot\Delta w^T\phi_n$$

Which can be further written as:

$$f\left((w+\Delta w)^T\phi_1, (w+\Delta w)^T\phi_2, \cdots, (w+\Delta w)^T\phi_N\right) = \left[\sum_{n=1}^{N}\frac{\partial f}{\partial(w^T\phi_n)}\cdot\phi_n^T\right]\Delta w$$

Note that here ϕ_n is short for $\phi(x_n)$. Based on the equation above, we can obtain:

$$\nabla_w f = \sum_{n=1}^{N}\frac{\partial f}{\partial(w^T\phi_n)}\cdot\phi_n^T$$

Now we focus on the derivative of function g with respect to w:

$$\nabla_w g = \frac{\partial g}{\partial(w^Tw)}\cdot 2w^T$$

In order to find the optimal w, we set the derivative of J with respect to w equal to 0, yielding:

$$\nabla_w J = \nabla_w f + \nabla_w g = \sum_{n=1}^{N}\frac{\partial f}{\partial(w^T\phi_n)}\cdot\phi_n^T + \frac{\partial g}{\partial(w^Tw)}\cdot 2w^T = 0$$

Rearranging the equation above, we can obtain:

$$w = \frac{1}{2a}\sum_{n=1}^{N}\frac{\partial f}{\partial(w^T\phi_n)}\cdot\phi_n$$

Where we have defined: $a = 1 \div \frac{\partial g}{\partial(w^Tw)}$, and since g is a monotonically increasing function, we have $a > 0$.

6.26

Using Eq(6.77), (6.78) and (6.86), we can obtain:

$$p(a_{N+1}|t_N) = \int p(a_{N+1}|a_N)\,p(a_{N+1}|t_N)\,da_N = \int N(a_{N+1}|k^TC_N^{-1}a_N, c-k^TC_N^{-1}k)\cdot N(a_N|a_N^*, H^{-1})\,da_N$$

By analogy to Eq(2.115), i.e.,

$$p(y) = \int p(y|x)\,p(x)\,dx$$

We can obtain:

$$p(a_{N+1}|t_N) = N(A\mu + b, L^{-1} + A\Lambda^{-1}A^T)$$

Where we have defined:

$$A = k^TC_N^{-1},\ b = 0,\ L^{-1} = c - k^TC_N^{-1}k \quad \text{And} \quad \mu = a_N^*,\ \Lambda = H$$

Therefore, the mean is given by:

$$A\mu + b = k^TC_N^{-1}a_N^* = k^TC_N^{-1}C_N(t_N - \sigma_N) = k^T(t_N - \sigma_N)$$

Where we have used Eq(6.84). The covariance matrix is given by:

$$L^{-1} + A\Lambda^{-1}A^T = c - k^TC_N^{-1} + k^TC_N^{-1}H^{-1}(k^TC_N^{-1})^T = c - k^T(C_N^{-1} - C_N^{-1}H^{-1}C_N^{-1})k$$
$$= c - k^T(C_N^{-1} - C_N^{-1}(W_N + C_N^{-1})^{-1}C_N^{-1})k = c - k^T(C_N^{-1} - (C_NW_NC_N + C_N^{-1})^{-1})k$$

Where we have used Eq(6.85) and the fact that C_N is symmetric. Then we use matrix identity (C.7) to further reduce the expression which will finally give Eq(6.88)