Bishop 5.3. $P(t|x,w) = \mathcal{N}(t|y(x,w),\Sigma)$

By 5.13 we know that. the likelihood function is:

$$P(t|x,w,\Sigma) = \prod_{n=1}^{N} P(t_n|x_n, w,\Sigma) \quad X = \{x_1,x_2,\cdots x_N\} \quad t=\{t_1,t_2\cdots t_N\}$$

$$= \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, w),\Sigma)$$

The log-likelihood function (K is the dimensionality of y and t):

$$\ln P(t_n|x_n, w,\Sigma) = -\frac{N}{2}(\ln|\Sigma| + K\ln(2\pi)) - \frac{1}{2}\sum_{n=1}^{N}(t_n - y(x_n,w))^T\Sigma^{-1}(t_n - y(x_n,w))$$

the error function:

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}(t_n - y_n)^T\Sigma^{-1}(t_n - y_n) \quad y_n = y(x_n,w)$$

maximizing the log-likelihood function w.r.t. Σ.

$$-\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(t_n - y_n)^T\Sigma^{-1}(t_n - y_n)$$

$$= -\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{Tr}\left[\Sigma^{-1}\sum_{n=1}^{N}(t_n - y_n)(t_n - y_n)^T\right]$$

maximize by setting the derivative with respect to $\Sigma^{-1} = 0$.

$$\Sigma = \frac{1}{N}\sum_{n=1}^{N}(t_n - y_n)(t_n - y_n)^T$$

Bishop 5.4 binary classification $t\in\{0,1\}$ output $y(x,w)$ that ~~pre~~ represents $P(t=1|x)$

set. real class label $k\in\{0,1\}$ we want $y(x,w) = p(k=1|x)$.

we know that: $p(t=1|x) = \sum_{k=0}^{1} P(t=1|k)(P(k|x)) = (1-\epsilon)y(x,w) + \epsilon(1-y(x,w))$

$$P(t|x) = p(t=1|x)^t(1-p(t=1|x))^{1-t}.$$

$$E(w) = -\sum_{n=1}^{N}\{t_n\ln[(1-\epsilon)y(x_n,w) + \epsilon(1-y(x_n,w))] + (1-t_n)\ln[1-(1-\epsilon)y(x_n,w) - \epsilon(1-y(x_n,w))]\}$$

we know that 5.21 $E(w) = -\sum_{n=1}^{N}\{t_n\ln y_n + (1-t_n)\ln(1-y_n)\}$ $\quad y_n = y(x_n,w)$

which is obtained by $\epsilon = 0$.

Bishop 5.26 $\Omega_n = \frac{1}{2}\sum_k (\mathcal{G}y_k)^2|_{x_n}$ $\quad \mathcal{G} \equiv \tau_i\frac{\partial}{\partial x_i}$. $\quad z_j = h(a_j)$ $\quad a_j = \sum_i w_{ji}z_i$

$\alpha_j = h'(a_j)\beta_j$ $\quad \beta_j = \sum_i w_{ji}\alpha_i$ where $\alpha_j \equiv \mathcal{G}z_j$ $\quad \beta_j \equiv \mathcal{G}a_j$

using $J_{ki} = \frac{\partial y_k}{\partial x_i}$ (5.70):

$$\Omega_n = \frac{1}{2}\sum_k\left(\sum_i \tau_{ni}\frac{\partial y_{nk}}{\partial x_{ni}}\right)^2 = \frac{1}{2}\sum_k\left(\sum_i \tau_{ni}J_{nki}\right)^2$$

summing this over n we can get. 5.128. $\Omega = \frac{1}{2}\sum_n\sum_k\left(\sum_{i=1}^{D}J_{nki}\tau_{ni}\right)^2$

we can see that β_{nl} can be written in terms of α_{ni}, which in turn can be written as functions of β_{ni} from previous layer. For the input layer. just use. α_i β_j

$$\beta_{nj} = \sum_i w_{ji}\alpha_{ni} = \sum_i w_{ji}\mathcal{G}x_{ni} = \sum_i w_{ji}\sum_{i'}\tau_{ni'}\frac{\partial x_{ni}}{\partial x_{ni'}} = \sum_i w_{ji}\tau_{ni}$$

$$\Omega_n = \frac{1}{2}\sum_k (\mathcal{G} y_{nk})^2 = \frac{1}{2}\sum_k \alpha_{nk}^2$$

using 5.52.

$$\frac{\partial \Omega_n}{\partial w_{rs}} = \sum_k (\mathcal{G} y_{nk}) \mathcal{G}(\delta_{nkr} z_{ns})$$

$$= \sum_k \alpha_{nk} (\phi_{nkr} z_{ns} + \delta_{nkr} \alpha_{ns})$$

using 5.74

$$\delta_{nkr} = h'(a_{nr}) \sum_l w_{lr} \delta_{nkl}$$

Back propagation equations for ϕ_{nkr}.

$$\phi_{nkr} = \mathcal{G}\delta_{nkr} = \mathcal{G}\left(h'(a_{nr}) \sum_l w_{lr} \delta_{nkl}\right)$$

$$= h''(a_{nr}) \beta_{nr} \sum_l w_{lr} \delta_{nkl} + h'(a_{nr}) \sum_l w_{lr} \phi_{nkl}.$$