Pour le réseau neuronnal de la figure 5.1, on a établi que pour une sortie y_k :

$$y_k = f(a_k) = f\left(\sum_{j=0}^{M} w_{kj}^{(2)} \sigma\left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i\right)\right)$$

Régression

Dans le cas d'une régression, il n'y a qu'une seule sortie :

$y(\vec{x}, \vec{w}) = f(a)$

Le modèle choisi :

$p(t\,|\vec{x}, \vec{w}) = \mathcal{N}(t\,|y(\vec{x}, \vec{w}), \beta^{-1})$

Puisque a peut approximer n'importe quelle fonction pour tout t, il n'est pas nécessaire de transformer a. On choisit :

$y(\vec{x}, \vec{w}) = f(a) = a$

$p(\vec{\mathbf{t}}\,|\boldsymbol{X}, \vec{w}, \beta) = \prod_{n=1}^{N} p(t_n|\vec{x}_n, \vec{w}, \beta)$

$E(\vec{w}) = \frac{1}{2} \sum_{n=1}^{N} \left\{y(\vec{x}_n, \vec{w}) - t_n\right\}^2$

Deux classes

Dans le cas d'une classification binaire, il n'y a qu'une seule sortie :

$y(\vec{x}, \vec{w}) = f(a)$

Le modèle choisi :

$p(t\,|\vec{x}, \vec{w}) = y(\vec{x}, \vec{w})^t \left\{1 - y(\vec{x}, \vec{w})\right\}^{1-t}$

Dans le modèle choisi, t suit une loi de Bernouilli de paramètre $y(\vec{x}, \vec{w})$. Cette loi est membre de la famille exponentielle. Lorsque la loi de Bernouilli est exprimée sous la forme générale de la famille exponentielle, la paramètre naturel η de la distribution est lié à l'espérance μ par la fonction ψ :

$\eta = \psi(\mu) = ln(\frac{\mu}{1-\mu})$

ψ est appelée la *fonction de liaison canonique*. Elle a un inverse :

$\mu = \psi^{-1}(\eta) = \sigma(\eta) = \frac{1}{1+exp(-\eta)}$

On choisit $f(a) = \psi^{-1}(a)$. En plus d'établir une correspondance entre les composantes du modèle et les paramètres de la distribution exponentielle, ce choix permet de tirer quelques généralisations additionnelles (voir section 4.3.6). On a donc :

$y(\vec{x}, \vec{w}) = f(a) = \sigma(a) = \frac{1}{1+exp(-a)}$

$p(\vec{\mathbf{t}}\,|\boldsymbol{X}, \vec{w}) = \prod_{n=1}^{N} y(\vec{x}_n, \vec{w})^{t_n} \left\{1 - y(\vec{x}_n, \vec{w})\right\}^{(1-t_n)}$

$E(\vec{w}) = -\sum_{n=1}^{N} t_n ln\left\{y(\vec{x}_n, \vec{w})\right\} + (1-t_n) ln\left\{1 - y(\vec{x}_n, \vec{w})\right\}$

Plusieurs classes

Dans le cas d'une classification multi-classes, il y K sorties :

$y_k(\vec{x}, \vec{w}) = f(a_k)$

Le modèle choisi :

$p(\vec{t}|\vec{x}, \vec{w}) = \prod_{k=1}^{K} \left\{y_k(\vec{x}, \vec{w})\right\}^{t_k}$

Dans le modèle choisi, $\vec{t}$ suit une loi multinomiale de paramètres K et $\vec{y}(\vec{x}, \vec{w})$. Cette loi est membre de la famille exponentielle. Lorsque la loi multinomiale est exprimée sous la forme générale de la famille exponentielle, la paramètre naturel $\vec{\eta}$ de la distribution est lié à l'espérance $\vec{\mu}$ par la fonction ψ :

$\eta_k = \psi(\mu_k) = ln(\frac{\mu_k}{1-\sum_j \mu_j})$

$\mu_k = \psi^{-1}(\eta_k) = \frac{exp(\eta_k)}{\sum_j exp(\eta_j)}$

On choisit $f(a_k) = \psi^{-1}(a_k)$:

$y_k(\vec{x}, \vec{w}) = f(a_k) = \frac{exp(a_k)}{\sum_j exp(a_j)}$

$p(\mathbf{T}\,|\boldsymbol{X}, \vec{w}) = \prod_{n=1}^{N} \prod_{k=1}^{K} y_k(\vec{x}_n, \vec{w})^{t_{nk}}$

où t_{nk} et la k^e composante de la cible $\vec{t}$ de la n^e observation.

$E(\vec{w}) = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk}\, ln\left\{y_k(\vec{x}_n, \vec{w})\right\}$