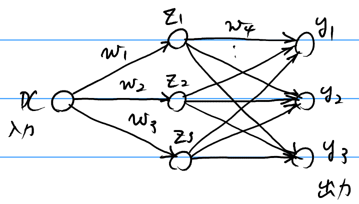
5.40

・ニューラルネットワークの出力ユニットの活性化関数をソフトマックス関数とする。
・目標変数 $\mathbf{t}$ の条件付分布を多項分布とし、ニューラルネットの出力をこの多項分布の平均と解釈する。
というモデルを考える



出力ユニット y_k の活性を a_k とすると

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

目標変数 $\mathbf{t}$ (1 of K表記) の条件付分布は

$$p(\mathbf{t} \mid x, w) = \prod_{k=1}^{K} y_k(x, w)^{t_k}$$

となる。このとき $D = \{\mathbf{t}_1, \mathbf{t}_2, \cdots\}$, $X = \{x_1, x_2, \cdots\}$, $y_{nk} = y_k(x_n, w)$ とおくと尤度は

$$p(D \mid w, X) = \prod_{n=1}^{N} p(\mathbf{t}_n \mid x_n, w) = \prod_{n=1}^{N} \prod_{k=1}^{K} y_{nk}^{t_{nk}}$$

対数尤度は

$$\ln p(D \mid w, X) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln y_{nk}$$

となる。
(パラメータ w の推定)
ベイズモデルなので w の推定値は最大事後確率推定(MAP推定)で与えられる。wの事後分布は

$$p(w \mid D, X) = \frac{p(D, w \mid X)}{p(D \mid X)} = \frac{p(D \mid w, X)\, p(w)}{p(D \mid X)} \propto p(D \mid w, X)\, p(w)$$

D, X は観測値に固定されているので $p(D \mid X)$ は定数である

なので対数事後分布は

$$\ln p(w \mid D, X) = \ln p(D \mid w, X) + \ln p(w) + C, \quad (C \text{は} w \text{によらない定数})$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln y_{nk} + \ln p(w) + C$$

となる。

wの事前分布を等方ガウス分布(5.162)とすると

$$\ln p(D|w,X) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk} - \frac{\alpha}{2} w^T w + C$$

となる。
よって、対数事後分布 $\ln p(D|w,X)$ の最大化は、(5.182)の正則化誤差関数

$$E(w) = -\ln p(D|w,X) + \frac{\alpha}{2} w^T w$$
$$= -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk} + \frac{\alpha}{2} w^T w$$

の最小化と等価であることが分かる。
$E(w)$ を最小にする w_{MAP} は $0 = \frac{\partial E}{\partial w}$ の解で与えられる。
この解は5.2.4節の勾配降下法で得ることができる。
勾配降下法で必要な $\frac{\partial E}{\partial w}$ の計算は5.3節の誤差逆伝播で行うことができる。

(予測分布)

新しい x に対する t の予測分布は

$$p(t|x,D,X) = \int p(t,w|x,D,X)\,dw = \int p(t|x,w)\,p(w|D,X)\,dw$$
$$= \int \prod_{k=1}^{K} y_k(x,w)^{t_k}\, p(w|D,X)\,dw$$

で与えられる。
ここに出てくる w の事後分布 $p(w|D,X)$ をラプラス近似(5.167)する。
ラプラス近似に必要なヘッセ行列 H は(5.91)若しくは(5.93),(5.94),(5.95)で得られる。

(超パラメータの推定)

超パラメータ α はエビデンス最大化によって(5.178)で与えられる。