5.5 ニューラルネットワークの正則化

入力と出力のユニットの数はデータの次元

隠れ層のユニットの数 M は任意　　M は ネットワークのパラメータの数を決める。

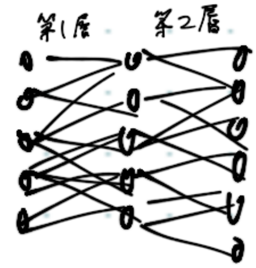
↑ 汎化性能を上げる良い M があるはず。

どうみつける？

・アプローチ1　いくつかMをためす。　図 5.10

・アプローチ2　Mを大きくとっておいて正則化項を追加 (5.112)

5.5.1 無矛盾なガウス事前分布



2層の重み、線形出力のNN

第1層 $z_j = h\left(\sum_i w_{ji} x_i + w_{j0}\right)$

第2層 $y_k = \sum_j w_{kj} z_j + w_{k0}$

入力を線形変換　$x_i \to \tilde{x}_i = a x_i + b$

NNによる写像を不変にするように重みとバイアスも変換

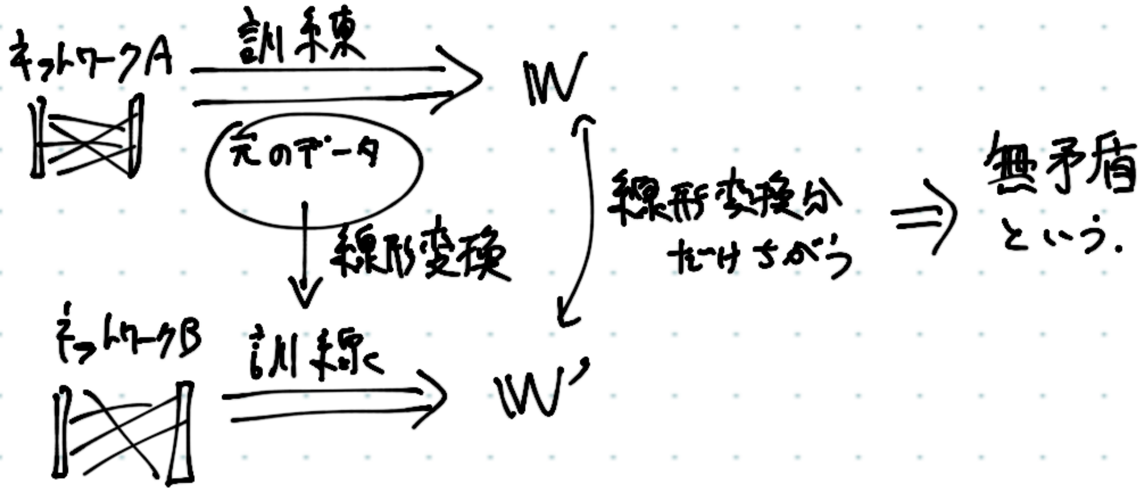
$w_{ji} \to \tilde{w}_{ji} = \frac{1}{a} w_{ji}$

$w_{j0} \to \tilde{w}_{j0} - \frac{b}{a} \sum_i w_{ji}$

同様に出力を $y_k \to \tilde{y}_k = c y_k + d$ とするときは

$w_{kj} \to \tilde{w}_{kj} = c w_{kj}$

$w_{k0} \to \tilde{w}_{k0} = c w_{k0} + d$



(5.112)の正則化項は無矛盾性をもたない。✗

上記の変換でも不変な正則化項を立ててる!

それがこれ!!

$$\boxed{\frac{\lambda_1}{2}\sum_{w\in W_1} w^2 + \frac{\lambda_2}{2}\sum_{w\in W_2} w^2 \quad (5.121)}$$

第1層　　第2層　　バイアスは入れてない。

$\lambda_1 \to a^{1/2}\lambda_1$, $\lambda_2 \to c^{-1/2}\lambda_2$ というリスケーリングで重みの変換の下で不変.

(5.121)は変則事前分布

$$p(w|\alpha_1,\alpha_2) \propto \exp\left(-\frac{\alpha_1}{2}\sum_{w\in W_1} w^2 - \frac{\alpha_2}{2}\sum_{w\in W_2} w^2\right) \quad (5.122)$$

に対応している.

$$p(w) \propto \exp\left(-\frac{1}{2}\sum_k \alpha_k \|w\|_k^2\right) \quad (5.123)$$

図5.11は に従って w を生成したNNの出力 入力 をプロット

5.6.2 早期終了