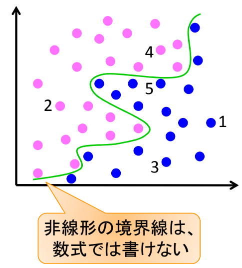
知能機械と自然言語処理
知能機械部　第12～13回

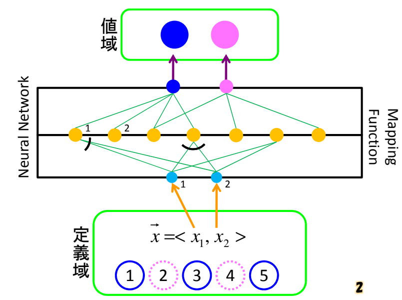
ソフトウェア情報学部

Goutam Chakraborty

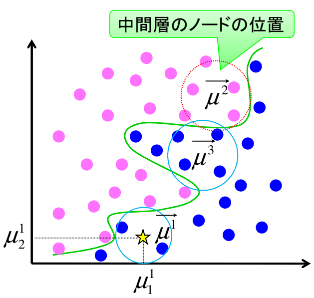
RBF（1）

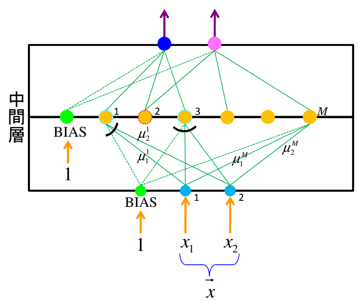
- Radical Basis Function Neural Network
  （放射基底関数ニューラルネットワーク）
  - 入力ノードから中間層ノードへの結合の重みを設定するため、データの教師なし分類を行う（クラスタリング）





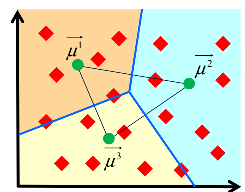
RBF（2）





RBF（3）

- 教師データ（例えば、にしん200匹，さんま200匹）からランダムに M 個データを選ぶ
  - $M \approx 400 \times 0.05 \approx 20$ （『≈』は、『≒』の国際的な記述方）
- クラスタリング
  - K-MEANS（教師なし学習）
  - Dirichlet Partition



- 中間層のノード数が少ないと、学習が出来ない（BIAS）
- 逆にノード数が多いと、テストデータは上手く分類されない（Variance）

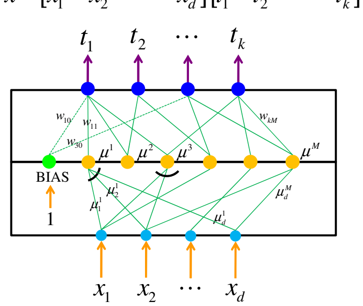
RBF（4）

2 Classの 2次元データ

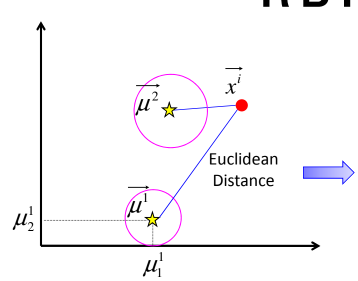
	データ		教師データ	
$\vec{x^1}$	x_1^1	x_2^1	t_1^1	t_2^1
$\vec{x^2}$	x_1^2	x_2^2	t_1^2	t_2^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vec{x^N}$	x_1^N	x_2^N	t_1^N	t_2^N

k Classの d 次元データ

$$\vec{x} = [x_1 \quad x_2 \quad \cdots \quad x_d][t_1 \quad t_2 \quad \cdots \quad t_k]$$



RBF（5）



$$\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

$$\Sigma = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$ の場合、上式は簡単になる！

$$\exp\left(-\frac{1}{2a^2}(|\vec{x}-\vec{\mu}|)^2\right)$$

距離： $(\vec{x}-\vec{\mu})^T(\vec{x}-\vec{\mu})$

RBF（6）

$$\varphi_M = \exp\left(-\frac{1}{2a^2}(|\vec{x}-\overrightarrow{\mu^M}|)^2\right)$$

ただし、 $\varphi_0 = 1$ (BIAS)

$$t_k = \varphi_0 w_{k0} + \varphi_1 w_{k1} + \cdots + \varphi_M w_{kM} \Rightarrow$$

2 Classの場合

$$\underbrace{\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}}_{2\times 1} = \underbrace{\begin{bmatrix} w_{10} & w_{11} & \cdots & w_{1M} \\ w_{20} & w_{21} & \cdots & w_{2M} \end{bmatrix}}_{2\times(M+1)} \underbrace{\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \vdots \\ \varphi_M \end{bmatrix}}_{(M+1)\times 1}$$

k Classの場合

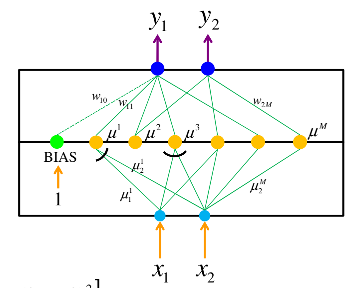
$$\underbrace{\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_k \end{bmatrix}}_{k\times 1} = \underbrace{\begin{bmatrix} w_{10} & w_{11} & \cdots & w_{1M} \\ w_{20} & w_{21} & \cdots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k0} & w_{k1} & \cdots & w_{kM} \end{bmatrix}}_{k\times(M+1)} \underbrace{\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \vdots \\ \varphi_M \end{bmatrix}}_{(M+1)\times 1}$$

RBF（7）

重みの値はどのように決める？ ← 教師データから学習する

1つのデータの場合

$$\text{誤差 } E = \frac{1}{2}[y-t]^2 = \frac{1}{2}\sum_{k=1}^{2}(y_k - t_k)^2$$



k	1	2
t_k^1	1	0
y_k	0.4	0.7

$$E = \frac{1}{2}\left[(0.4-1)^2 + (0.7-0)^2\right] = \frac{1}{2}(0.36+0.49) = 0.425$$

RBF（8）

全てのデータに対して、誤差を求める

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k}(y_j^i - t_j^i)^2 \Rightarrow E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k}(W\Phi - t_j)^2$$

$$\begin{bmatrix} y_1^1 \\ y_2^1 \end{bmatrix} = W\begin{bmatrix} \varphi_0^1 & \varphi_1^1 & \cdots & \varphi_M^1 \end{bmatrix}^T$$

$$\begin{bmatrix} y_1^2 \\ y_2^2 \end{bmatrix} = W\begin{bmatrix} \varphi_0^2 & \varphi_1^2 & \cdots & \varphi_M^2 \end{bmatrix}^T$$

$$\vdots \qquad\qquad \vdots$$

$$\begin{bmatrix} y_1^N \\ y_2^N \end{bmatrix} = W\begin{bmatrix} \varphi_0^N & \varphi_1^N & \cdots & \varphi_M^N \end{bmatrix}^T$$

E が最小になる値を見つけるには、

微分 $\dfrac{dE}{dW} = 0$ の解から

W を求めることができる。

RBF（9）

$$\boxed{\Phi^T \Phi W^T = \Phi^T \mathrm{T}}$$

Psuedo Inverse

$$W^T = \underbrace{\underbrace{(\underbrace{\underbrace{\Phi^T}_{M\times N}\underbrace{\Phi}_{N\times M}}_{M\times M})^{-1}\underbrace{\Phi}_{M\times N}}_{M\times N}\underbrace{\mathrm{T}}_{N\times k}}_{M\times k}$$

$$\mathrm{T} \quad = \quad \Phi \quad \times \quad W^T$$

$$\begin{bmatrix} y_1^1 & y_2^1 \\ y_1^2 & y_2^2 \\ & \\ y_1^k & y_2^k \end{bmatrix} = \underbrace{\begin{bmatrix} [\varphi_0^1 & \varphi_2^1 & \cdots & \varphi_M^1] \\ [\varphi_0^2 & \varphi_2^2 & \cdots & \varphi_M^2] \\ \vdots & \vdots & \ddots & \vdots \\ [\varphi_0^N & \varphi_2^N & \cdots & \varphi_M^N] \end{bmatrix}}_{N\times(M+1)} \begin{matrix} W^T \\ W^T \\ \vdots \\ W^T \end{matrix}$$

$$\therefore\ \Phi^{-1} = (\Phi^T\Phi)^{-1}\Phi$$

$$\mathrm{T} = \Phi W^T$$

$$\Phi \cancel{\mathrm{T} =} W^T$$

長方行列の逆行列は求められない

自由課題

- にしん200, さんま200のデータを使って、RBFを求めよ。