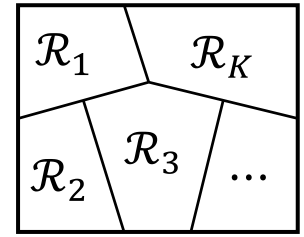
平均誤り確率

決定則

- 入力 $\boldsymbol{x}$ に対してそれが属すべきクラスを与える規則：$\alpha(\boldsymbol{x})$
- クラス数をKとすると、決定則により入力空間は互いに重ならないK個の**決定領域** $\mathcal{R}_k$（クラスω_kに判定する領域）に分割される
- 決定領域の境界を**決定境界**(あるいは**識別境界**)と呼ぶ

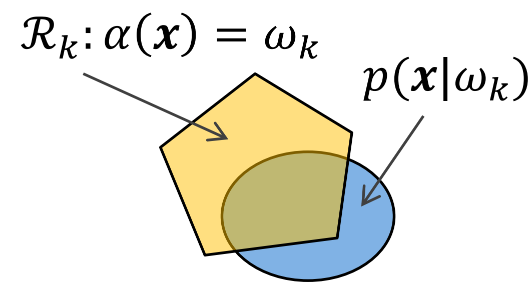


平均誤り確率

- クラス ω_k に属する入力 $\boldsymbol{x}$ を ω_k 以外に誤る確率

$$P_{err}(\omega_k) = \int_{\overline{\mathcal{R}_k}} p(\boldsymbol{x}|\omega_k)d\boldsymbol{x} = 1 - \int_{\mathcal{R}_k} p(\boldsymbol{x}|\omega_k)d\boldsymbol{x}$$

$\mathcal{R}_k$以外の領域



- 平均誤り確率

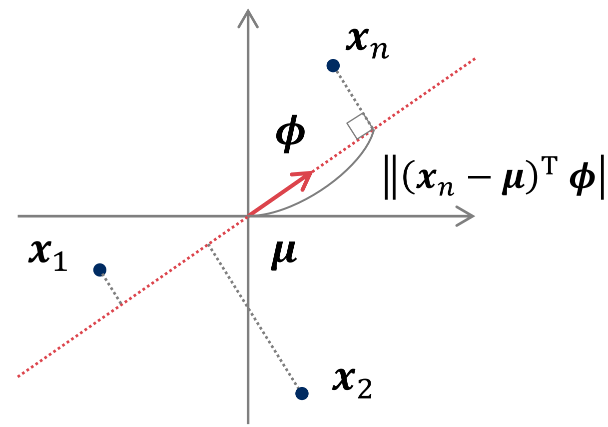
$$P_{err} = \sum_{k=1}^{K} P_{err}(\omega_k)P(\omega_k) = \sum_{k=1}^{K} P(\omega_k) - \sum_{k=1}^{K}\int_{\mathcal{R}_k} p(\boldsymbol{x}|\omega_k)P(\omega_k)d\boldsymbol{x}$$

$$= 1 - \sum_{k=1}^{K}\int_{\mathcal{R}_k} p(\boldsymbol{x}|\omega_k)P(\omega_k)d\boldsymbol{x} \underset{\text{ベイズの定理}}{=} 1 - \sum_{k=1}^{K}\int_{\mathcal{R}_k} p(\boldsymbol{x})\underline{P(\omega_k|\boldsymbol{x})}\, d\boldsymbol{x}$$

- 積分内部が最大になるように $\mathcal{R}_k$を設定すれば、平均誤り確率は最小化される
  ⇒ $p(\boldsymbol{x})$ はクラスに無関係なので、$\mathcal{R}_k$内で $\underline{P(\omega_k|\boldsymbol{x})}$ が最大となるように設定すればよい

共分散行列と主成分分析

主成分分析は共分散行列の固有値問題に等しい



- ベクトル $\boldsymbol{\phi}$ に投影した時の分散

$$V = \frac{1}{N}\sum_{n=1}^{N}\left\|(\boldsymbol{x}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\phi}\right\|^2$$

$$= \left\|\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi}\right\|^2$$

$$\mathrm{X} = \frac{1}{\sqrt{N}}(\boldsymbol{x}_1 - \boldsymbol{\mu}, \cdots, \boldsymbol{x}_N - \boldsymbol{\mu})$$

- 分散を最大化する単位ベクトル（ラグランジュ未定乗数法を用いる）

$$L = \left\|\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi}\right\|^2 - \lambda(\|\boldsymbol{\phi}\|^2 - 1) = \boldsymbol{\phi}^{\mathrm{T}}\mathrm{X}\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi} - \lambda\left(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi} - 1\right)$$

$$\frac{\partial L}{\partial \boldsymbol{\phi}} = 2\mathrm{X}\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi} - 2\lambda\boldsymbol{\phi} = 0 \quad \longleftarrow \quad \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\mathrm{A}\boldsymbol{x}) = \left(\mathrm{A} + \mathrm{A}^{\mathrm{T}}\right)\boldsymbol{x}$$

$$\rightarrow \quad \mathrm{X}\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}$$

$$\rightarrow \quad \Sigma\boldsymbol{\phi} = \lambda\boldsymbol{\phi} \quad \longleftarrow \quad \Sigma = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^{\mathrm{T}} = \mathrm{X}\mathrm{X}^{\mathrm{T}}$$

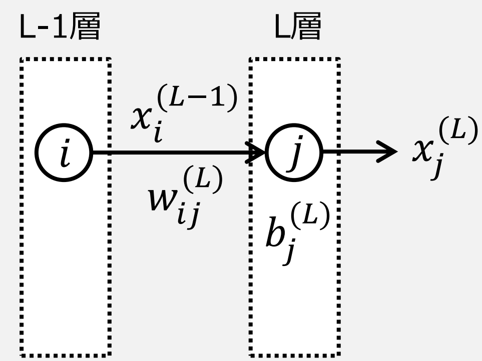
共分散行列の固有値問題

固有値は分散に等しい：$V = \left\|\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi}\right\|^2 = \boldsymbol{\phi}^{\mathrm{T}}(\mathrm{X}\mathrm{X}^{\mathrm{T}}\boldsymbol{\phi}) = \boldsymbol{\phi}^{\mathrm{T}}(\lambda\boldsymbol{\phi}) = \lambda$

Multi-Layer Perceptron (MLP) の学習

誤差逆伝播学習 (Back propagation)

$$E = \frac{1}{2}\sum_j \left(x_j^{(L_{max})} - t_j\right)^2$$ （オンライン学習の目的関数）



$$u_j^{(L)} = \sum_i w_{ij}^{(L)} x_i^{(L-1)} + b_j^{(L)}$$

$$x_j^{(L)} = f(u_j^{(L)})$$

<u>最急降下法</u>

$$w_{ij}^{(L)} \leftarrow w_{ij}^{(L)} - \eta \frac{\partial E}{\partial w_{ij}^{(L)}} \qquad \frac{\partial E}{\partial w_{ij}^{(L)}} = \frac{\partial E}{\partial u_j^{(L)}} \frac{\partial u_j^{(L)}}{\partial w_{ij}^{(L)}} = \frac{\partial E}{\partial u_j^{(L)}} x_i^{(L-1)}$$

$L = L_{max}$ の時

$$\frac{\partial E}{\partial u_j^{(L)}} = \left(x_j^{(L)} - t_j\right) \frac{\partial x_j^{(L)}}{\partial u_j^{(L)}} = \left(x_j^{(L)} - t_j\right) f'\left(u_j^{(L)}\right) \equiv \delta_j^{(L)}$$

$$\frac{\partial E}{\partial u_j^{(L-1)}} = \sum_i \frac{\partial E}{\partial u_i^{(L)}} \frac{\partial u_i^{(L)}}{\partial u_j^{(L-1)}} = \sum_i \delta_i^{(L)} w_{ji}^{(L)} f'(u_j^{(L-1)}) \equiv \delta_j^{(L-1)} \qquad u_i^{(L)} = \sum_j w_{ji}^{(L)} f\left(u_j^{(L-1)}\right) + b_i^{(L)}$$

$$\boxed{w_{ij}^{(L)} \leftarrow w_{ij}^{(L)} - \eta \delta_j^{(L)} x_i^{(L-1)}}$$

$$\boxed{b_j^{(L)} \leftarrow b_j^{(L)} - \eta \delta_j^{(L)}}$$

$$\delta_j^{(L)} = \begin{cases} \left(x_j^{(L)} - t_j\right) f'\left(u_j^{(L)}\right) & (L = L_{max}) \\ \sum_i \delta_i^{(L+1)} w_{ji}^{(L+1)} f'(u_j^{(L)}) & (L \neq L_{max}) \end{cases}$$

誤差 $\delta_j^{(L)}$ が下の層に伝播する ⇒ **誤差逆伝播**

Support Vector Machine (3/6)

双対問題

$$L_P = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N} r_n\,\xi_n - \sum_{n=1}^{N}\alpha_n\{(\boldsymbol{w}\cdot\boldsymbol{x}_n - b)t_n - 1 + \xi_n\}$$

$$\frac{\partial L_P}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{n=1}^{N}\alpha_n t_n \boldsymbol{x}_n = 0, \quad \frac{\partial L_P}{\partial b} = \sum_{n=1}^{N}\alpha_n t_n = 0, \quad \frac{\partial L_P}{\partial \xi_n} = C - r_n - \alpha_n = 0$$

$r_n \geq 0$ より $\alpha_n \leq C$

これらを L_P に代入すると、 $\boldsymbol{\alpha}$ に関する最適化に書き換えられる ($L_P \rightarrow L_D$)

$$\frac{1}{2}\|\boldsymbol{w}\|^2 = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j t_i t_j \boldsymbol{x}_i\cdot\boldsymbol{x}_j\,, \qquad C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N} r_n\,\xi_n = \sum_{i=1}^{N}\alpha_i\,\xi_i$$

$$\sum_{n=1}^{N}\alpha_n\{(\boldsymbol{w}\cdot\boldsymbol{x}_n - b)t_n - 1 + \xi_n\} = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j t_i t_j \boldsymbol{x}_i\cdot\boldsymbol{x}_j - b\underbrace{\sum_{i=1}^{N}\alpha_i t_i}_{=0} - \sum_{i=1}^{N}\alpha_i + \sum_{i=1}^{N}\alpha_i\xi_i$$

$$\underset{\boldsymbol{\alpha}}{maximize} \quad L_D = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j t_i t_j \underbrace{\boldsymbol{x}_i\cdot\boldsymbol{x}_j}_{\rightarrow K(\boldsymbol{x}_i,\boldsymbol{x}_j)}$$

$$subject\ to \quad \sum_{i=1}^{N}\alpha_i t_i = 0, \qquad 0 \leq \alpha_i \leq C \quad (i = 1,\cdots,N)$$

上に凸な2次関数