ニールソン

Learning Machine

Nils. J. Nilsson

C. 2

2.2 discriminant function

$$g(x) = w_1 x_1 + \cdots + w_d x_d + w_{d+1}$$

2.3 minimum distance classifiers

$$|x - P_i|^2 = X \cdot X - 2X \cdot P_i + P_i \cdot P_i$$

$$g_i(x) = X \cdot P_i - \frac{1}{2} P_i \cdot P_i$$

$$P_{ij} = w_{ij} \,, \quad w_{i \cdot d+1} = -\frac{1}{2} P_i \cdot P_i$$

2.4 the dicision surface of linear machine

$$g_i(x) - g_j(x) = 0$$

the dicision regions of a linear machine are always convex.

2.5 linear classifications of patterns

a finite set $\mathcal{X}$ of distinct patterns $\{X_1, X_2 \cdots X_N\}$

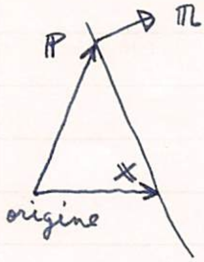
$\mathcal{X} \rightarrow$ subsets $\mathcal{X}_1$ $\mathcal{X}_2$ $\cdots$ $\mathcal{X}_R$

$$\boxed{\begin{array}{l} g_i(x) > g_j(x) \text{ for all } X \text{ in } \mathcal{X}_i \\ j = 1 \cdots R. \quad j \neq i \quad \text{ for all } i = 1 \cdots R \end{array}}$$

$\Downarrow$ a classification of $\mathcal{X}$ is linear

subsets $\mathcal{X}_1 \cdots \mathcal{X}_R$ are linearly separable

2.6 TLU the threshold logic unit



$$g(\mathbb{X}) = w_1 x_1 + \cdots + w_d x_d + w_{d+1}$$

$$\begin{cases} \mathbb{X}\cdot\mathbb{W} = -w_{d+1} \\ \mathbb{X}\cdot n = \mathbb{P}\cdot n \end{cases}$$

$$n = \frac{\mathbb{W}}{|\mathbb{W}|} \qquad n\cdot\mathbb{P} = -\frac{w_{d+1}}{|\mathbb{W}|} = -\delta_w$$

hyperplane $\mathbb{X}\cdot n + \delta_w = 0$

2.7 Piecewise linear discriminant functions

R finite point sets $\mathcal{P}_1 \cdots \mathcal{P}_R$

$\mathcal{P}_i \xrightarrow{\in} \mathbb{P}_i^{(1)} \cdots \mathbb{P}_i^{(L_i)}$

$$d(\mathbb{X}, \mathcal{P}_i) = \min_{j=1\cdots L_i} |\mathbb{X} - \mathbb{P}_i^{(j)}|$$

$$g_i(\mathbb{X}) = \max\{\mathbb{P}_i^{(j)}\mathbb{X} - \tfrac{1}{2}\mathbb{P}_i^{(j)}\cdot\mathbb{P}_i^{(j)}\} = \max\{g_i^{(j)}(\mathbb{X})\}$$

ここで

subsidiary discriminant function として

$$g_i^{(j)}(\mathbb{X}) = w_{i1}^{(j)} x_1 \cdots w_{id}^{(j)} x_d + w_{i,d+1}^{(j)}$$

また

$\mathbb{P}_i^{(j)}(p_{i1}^{(j)}, \cdots, p_{id}^{(j)})$ とすると

$$w_{ik}^{(j)} = p_{ik}^{(j)} \quad \begin{cases} i = 1 \cdots R \\ j = 1 \cdots L_i \\ k = 1 \cdots d \end{cases}$$

$$w_{i,d+1}^{(j)} = -\tfrac{1}{2}\mathbb{P}_i^{(j)}\cdot\mathbb{P}_i^{(j)}$$

<u>minimum distance classfier の discriminant function</u>

として piecewise linear discriminant function

2.8 quadric discriminant functions

$$g_i(\mathbb{X}) = \sum_{j=1}^{d} w_{jj} x_j^2 + \sum_{j=1}^{d-1} \sum_{k=j+1}^{d} w_{jk} x_j x_k + \sum_{j=1}^{d} w_j x_j + w_{d+1}$$

$A = [a_{jk}]$ $\quad a_{jj} = w_{jj} \quad j = 1 \cdots d$

$\quad a_{jk} = \frac{1}{2} w_{jk} \quad j,k = 1 \cdots d,\ j \neq k$

$\mathbb{B} = \begin{pmatrix} b_1 \\ \vdots \\ b_d \end{pmatrix}$ $\quad b_j = w_j \quad j = 1 \cdots d$

$C = w_{d+1}$

$$g(\mathbb{X}) = \underbrace{\mathbb{X}^t \mathbb{A} \mathbb{X}}_{\text{quadratic form}} + \mathbb{X}^t \mathbb{B} + \mathbb{C}$$

~~gu~~ quadratic form

all the eigenvalues of $\mathbb{A} > 0$ → quadratic form $\geqq 0$

A and quadratic form is positive definite

the eigenvalues of $\mathbb{A} > 0$ or $= 0$

$\mathbb{A}$ and quadratic form is positive semidefinite

2.9. quadric decision surfaces

R_i と R_j との境界

$$\mathbb{X}^t [\mathbb{A}^{(i)} - \mathbb{A}^{(j)}] \mathbb{X} + \mathbb{X}^t [\mathbb{B}^{(i)} - \mathbb{B}^{(j)}] + [C^{(i)} - C^{(j)}] = 0$$

$$\mathbb{X}^t \mathbb{A} \mathbb{X} + \mathbb{X}^t \mathbb{B} + C = 0$$

$\mathbb{A}$ is positive definite → hyperellipsoid

the axes are in the direction of the eigenvalue of $\mathbb{A}$

A is an identity matrix → hypersphere

A is positive semidefinite → hyperellipsoidal cylinder

上記以外の $\mathbb{A}$ のとき → ~~hyper~~hyperboloid

2.10 Implementation of quadric discriminant function

$$g_i(\mathbb{X}) = \sum_{j=1}^{d} w_{jj} x_j^2 + \sum_{j=1}^{d-1} \sum_{k=j+1}^{d} w_{jk} x_j x_k + \sum_{j=1}^{d} w_j x_j + w_{d+1}$$

$$g(\mathbb{X}) = w_1 f_1 + \cdots + w_M f_M + w_{M+1}$$

$$\mathbb{F} = \mathbb{F}(\mathbb{X}) = \mathbb{F}(f_1, f_2, \cdots f_M)$$

$$\mathbb{X} = \mathbb{X}(x_1, x_2 \cdots x_d).$$

$\mathbb{X}$のすべての2次識別関数は$\mathbb{F}$の線形識別関数に対応させうる

2.11. Φ functions

この式は weights $w_1\ w_2 \cdots w_M\ w_{M+1}$ に関して linear

∴ Φ function

$f_i(\mathbb{X})$ は weight に関して独立

線形独立、実、一価関数

$$\mathbb{X} \xrightarrow[\text{写像}]{\mathbb{F}(\mathbb{X})} \mathbb{F} \quad (\underbrace{M\text{次元空間}}_{\Phi\text{-space}})$$

Φ関数を用いたパターン分類機 = Φ-machine

= Φ processor + linear machine

piecewise linear machine = Φ machine

2.12 the utility of Φ functions for classifying patterns.

one measure of the effectiveness

the total number of dichotomies of N patterns

2.13 the number of linear dichotomies of N-points of d-dimentional

$L(N, d)$ for N d-dimentional patterns
the number of linear dichotomies

general position

$N > d$ no subset of $d+1$ points lies on a $(d-1)$-dimentional hyperplane

$N \leq d$ no $(N-2)$-dimentional hyperplane contains the set of N points

$$L(N, d) = L(N-1, d) + L_{X_N}(N-1, d)$$

$$L_{X_N}(N-1, d) = L(N-1, d-1)$$

$$\therefore L(N, d) = L(N-1, d) + L(N-1, d-1).$$

また $L(1, d) = 2$. $L(N, 1) = 2N$

$$\therefore L(N, d) = 2 \sum_{i=0}^{d} \binom{N-1}{i} \quad \text{for } N > d$$

$$= 2^N \quad \text{for } N \leq d$$

2.14 the effects of constraints

a set $\mathcal{X}$ of N points

a set $\mathcal{Z}$ of K points $\quad K < d$

$L_{\mathcal{Z}}(N, d)$ = the number of linear dichotomies of $\mathcal{X}$ achievable by a hyperplane constrained to contain all the points of $\mathcal{Z}$.

$$L_{\mathcal{Z}}(N, d) = L(N, d-K)$$

2.15 the number of Φ function dichotomies

Φ function family $g(x) = w_1 f_1(\mathbf{x}) + \cdots + w_M f_M(\mathbf{x}) + w_{M+1}$

$g(x)=0$ による N点集合 $\mathcal{X}$ の分割法 "Φ二分法"

パターンspaceに $M+1$ ケ以上の $\mathcal{X}$ のメンバを含む $g(x)=0$ はないとき

"Φ一般位置"

$$\Phi(N, d) = L(N, M)$$

1. $\Phi(\mathbf{x}) = |\mathbf{x} - \mathbf{w}|^2 - a^2 = \mathbf{x}\cdot\mathbf{x} - 2\mathbf{x}\cdot\mathbf{w} + \mathbf{w}\cdot\mathbf{w} - a^2$

   $M = d+1$ (a hypersphere decision surface)

2. $\Phi(\mathbf{x})$: r 次多項関数.

$$M = \sum_{i=1}^{r} \binom{d+i-1}{i} = \binom{d+r}{r} - 1$$

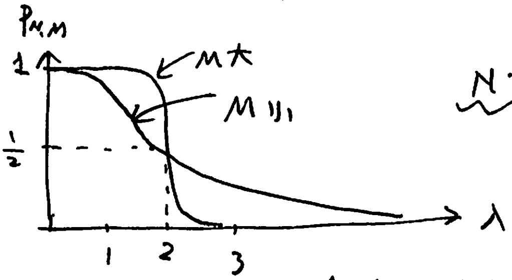
$r=2$ のとき $M = \dfrac{d(d+3)}{2}$

(the general quadric surface)

2.6 machine capacity

$P_{N,M}$; N patterns の可能な2分法の数 2^N 個;
任意に1つ選んだとき、与えられた Φ machine で
重みを適当に選んで道具立てできる probability

$$P_{N,M} = \frac{L(N,M)}{2^N} = \begin{cases} 2^{1-N} \sum_{i=0}^{M} \binom{N-1}{i} & N > M \\ = 1 & N < M \end{cases}$$



$\underline{N \simeq \lambda(M+1)}$

M大で threshold effect

Capacity C of a Φ machine $= 2(M+1)$

~~Φ machine で implemented される pattern space の d~~

Decision boundary in pattern space implemented by Φ machine	Capacity
Hyperplane	$2(d+1)$
Hypersphere	$2(d+2)$
General quadric surface	$(d+1)(d+2)$
rth order polynominal surface	$2\binom{d+r}{r}$

Chapter 3 parametric training methods

3.1 probabilistic pattern sets

the parametric Training method

1. $P(X|i)$ のパラメータ値とパラメータ $P(i)$ の値を用いて識別関数を表す
2. 〃 を訓練パターン集合を用いて推定する.
3. この推定値をパラメータの真の値とみなして 1. の識別関数の表示に用いる.

3.2 discriminant functions based on dicision theory

loss function $\lambda(i/j)$

the average value of $\lambda(i/j)$ is minimized → optimum

conditional average loss

$$L_x(i) = \sum_{j=1}^{R} \lambda(i|j)\, P(j|X)$$

the optimum machine makes its classification.

1. pattern X が現れる
2. $L_x(i)$ を計算 $i = 1 \cdots R$
3. $L_x(i_0) \leq L_x(i)$ i_0 をカテゴリ番号にする.

よって. $L_x(i)$ に負号をつければ識別関数の集合になる。

3.3. Likelihoods

$$P(j|X) = \frac{P(X|j)\, P(j)}{P(X)}$$

$$L_x(i) = \frac{1}{P(X)} \sum_{j=1}^{R} \lambda(i|j)\, P(X|j)\, P(j) \qquad \text{minimize}$$

$$\downarrow$$

$$l_x(i) \triangleq \sum_{j=1}^{R} \lambda(i|j)\, P(X|j)\, P(j) \qquad \text{minimize}$$

3.4 a special loss function

$\lambda(i/j) = 1 - \delta_{ij}$: symmetrical loss function

$$\underbrace{l_x(i)}_{\text{minimum}} = P(x) - \underbrace{P(x|i)\,P(i)}_{\text{maximum}}$$

すべてのカテゴリの事前生起確率が等しい $P(i) = \frac{1}{R}$ とき
$P(x|i)$ を maximum にすることを求めれば良い..
maximum likelihood decision
最尤決定

$\lambda(i/j) = 1 - \delta_{ij}$ のときは
識別関数は $g_i(x) = P(x|i)\,P(i)$
or
$g_i(x) = \log(x|i) + \log P(i)$

3.5 An example.

d 個の2値成分を持つパターンを分類 $x_i = 0$ or 1
$R = 2$

∘ $\lambda(i/j)$: symmetrical loss function とすると

識別関数 $g(x) = g_1(x) - g_2(x) = \log\left[\frac{P(x|1)}{P(x|2)}\right] + \log\left[\frac{P(1)}{1-P(1)}\right]$

∘いずれのカテゴリでも x の各成分は統計的に独立にすると
$P(x|j) = P(x_1|j) \cdots P(x_d|j)$
また $P(x_i=1|1) \triangleq p_i$
$P(x_i=1|2) \triangleq q_i$) に表すと

$$g(x) = \sum_{i=1}^{d} \log\left[\frac{P(x_i|1)}{P(x_i|2)}\right] + \log\left[\frac{P(1)}{1-P(1)}\right] = \sum_{i=1}^{d} x_i \log\left[\frac{p_i(1-q_i)}{q_i(1-p_i)}\right]$$
$$+ \sum \log\left(\frac{1-p_i}{1-q_i}\right) + \log\left[\frac{P(1)}{1-P(1)}\right]$$

3.6 the biviriate normal probability-density function

the expectation operator $E[\ ]$

$E[x_1] = m_1, \quad E[x_2] = m_2$: mean value

$E[x_1^2] - E^2[x_1] = \sigma_{11}, \quad E[x_2^2] - E^2[x_2] = \sigma_{22}$: variance

$$z_1 = \frac{x_1 - m_1}{\sqrt{\sigma_{11}}} \qquad z_2 = \frac{x_2 - m_2}{\sqrt{\sigma_{22}}}$$

$E[z_1] = 0 \quad E[z_1^2] - E^2[z_1] = 1.$

2変数正規密度関数

$$p(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\sigma_{12}^2}} \exp\left\{-\frac{1}{2}\frac{z_1^2 - 2\sigma_{12}z_1z_2 + z_2^2}{1-\sigma_{12}^2}\right\}$$

$\sigma_{12} = E[z_1 z_2]$: covariance or corelation

z_1, z_2 平面で
等確率密度を示す等高線はだ円

仮定 2次元パターンが2変数正規分布から選択できるとする
: 2変数正規パターン
∴ パターン点は (m_1, m_2) を中心とするだ円状集積

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} . \ M = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$: covariance matrix

$$p(X) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(X-M)^t \Sigma^{-1} (X-M)\right\}$$

3.7 the multivariate normal distribution

↓

d成分結合確率密度を表すことが可

これに従うパターン … 多変数正規パターン

d成分正規~~確率~~分布の式

$$P(\mathbb{X}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbb{X}-M)^t\Sigma^{-1}(\mathbb{X}-M)\right\}$$

$$\Sigma = E[(\mathbb{X}-M)(\mathbb{X}-M)^t] \quad , \quad M = E[\mathbb{X}]$$

covariance matrix　　　　mean vector

~~新~~略記

$$P(\mathbb{X}) \sim N(M, \Sigma).$$

3.8. the optimum classifier for normal patterns

仮定　事前確率 $P(i)$ 及び d変数正規確立関数 $P(\mathbb{X}|i)$ が既知とする

$$P(\mathbb{X}|i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(\mathbb{X}-M_i)^t\Sigma_i^{-1}(\mathbb{X}-M_i)\right\}$$

仮定　symmetrical loss function のとき

$$g_i(X) = \log P(\mathbb{X}|i) + \log P_i \qquad P_i = P(i)$$

結局 $g_i(\mathbb{X}) = b_i - \frac{1}{2}[(\mathbb{X}-M_i)^t\Sigma_i^{-1}(\mathbb{X}-M_i)]$

$b_i = \log P_i - \frac{1}{2}\log|\Sigma_i|$　また $-\frac{d}{2}\log 2\pi$ は無視

3.9 some special cases involving identical covariance matrix

仮定 $\Sigma_i = \Sigma$: iによらず一定

iによらない項を略すると

3.8の discriminant function は、

$$g_i(X) = X^t \Sigma^{-1} M_i + \log P_i - \frac{1}{2} M_i^t \Sigma^{-1} M_i$$

(線形識別関数となる)

例 R=2.

$$g(X) = X^t \Sigma^{-1}(M_1 - M_2) - \frac{1}{2} M_1^t \Sigma^{-1} M_1 + \frac{1}{2} M_2^t \Sigma^{-1} M_2 + \log \frac{P_1}{P_2}$$

$g(X) = 0$ は $\Sigma^{-1} M_1$ と $\Sigma^{-1} M_2$ の垂直2等分線

追加仮定 $\Sigma = I$ (単位行列) and $P_i = \frac{1}{R}$

$$g_i(X) = X \cdot M_i - \frac{1}{2} M_i \cdot M_i$$

(2章の点に関する最小きょり分類器と同じ)

3.10 training with normal pattern sets

正規パターンに対する最適分類機械 すなわち 二次機械の形は分布の形だけによる。

仮定 R個の category に属する代表パターンの訓練集合があるとする

sample statistics $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_R$

$$\langle X \rangle_i = \frac{1}{N_i} \sum_{X \in \mathcal{X}_i} X \quad ; \text{ sample mean}$$

$$\langle \Sigma \rangle_i = \frac{1}{N_i} \sum_{X \in \mathcal{X}_i} (X - \langle X \rangle_i)(X - \langle X \rangle_i)^t \quad ; \text{ sample covariance matrix}$$

この時 M_i, Σ_i の推定値 → 識別関数 : パラメータ訓練法

Q_i = X_iの各 pattern が列となる matrix

$$\langle \Sigma \rangle_i = \frac{1}{N_i} Q_i Q_i^t$$

仮定 以上は Σ_iは正則としてきた

Σ_iが nonsingler $\longrightarrow N_i \geq d$のとき $\langle \Sigma \rangle_i$も singular

3.11. learning the mean vector of normal patterns

仮定

共分散行列はすべて既知，平均ベクトルはランダム変数

仮定 pattern vectorは上の仮定をみたす正規分布とする．

既知の任意の Mに対する Xの分布

$$P(X \mid M) \sim N(M, \Sigma)$$

仮定．$P(M) \sim N(\mu, K)$

ここで $X = \Xi + M$ とする　Ξ：平均 0，共分散行列 Σ

$\therefore P(X) \sim N(\mu, \Sigma + K)$

Chapter 4

4.1 TLUのノン・パラメータ形訓練

•仮定• 訓練部分集合は線形分離可能

この章での訓練法

重みの調整を繰返して行う。すなわち、hyperplaneの位置と方向を繰返して変化させる。

4.2 重み空間

$Y = (x_1, \cdots, x_d, 1)$ 付加パターンベクトル

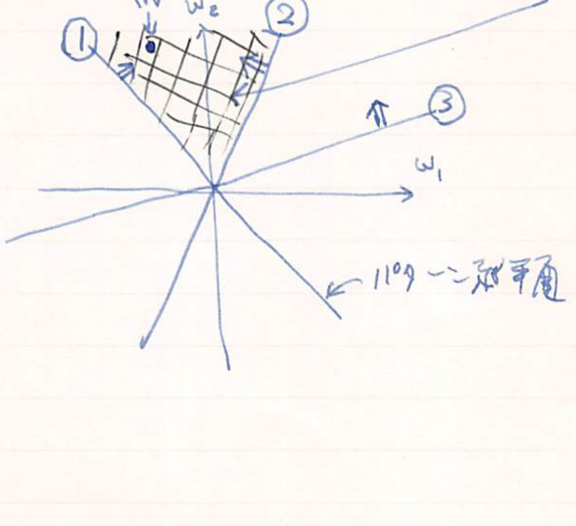
Xの線形識別関数 $g(X) = Y \cdot W$

$W \cdot Y = 0$: パターン超平面

解重みベクトル W

$$\begin{cases} Y \cdot W > 0 & \mathcal{Y}_1 \text{ に属する } Y \\ Y \cdot W < 0 & \mathcal{Y}_2 \quad 〃 \end{cases}$$

この解重み点 W のすべてを含む : 解地域



4.3 TLU training procedures

$$\mathbb{W}' = \mathbb{W} + c\,\mathbb{Y} \qquad c：訂正増分$$

注. 重み点をパターン超平面に垂直に移動する手順に限る.
誤り訂正訓練手順
1. 固定増分法
c ; positive constant
2. 絶対訂正法
$$c = \min\left\{ x \,;\, x > \frac{|\mathbb{W}\cdot\mathbb{Y}|}{\mathbb{Y}\cdot\mathbb{Y}} \text{ and } x : \text{integer} \right\}$$

3. 倍数訂正法
$$c = \lambda \frac{|\mathbb{W}\cdot\mathbb{Y}|}{\mathbb{Y}\cdot\mathbb{Y}} \quad , \lambda : \text{positive}$$

1、2、は有限回の訂正で解重みベクトルに至る：収束性

4.4 a numerical example of error correction training

4.5 an error-correction training procedure for $R > 2$.

仮定 訓練パターン集合 $\mathcal{Y}$ は $\mathcal{Y}_1, \cdots \mathcal{Y}_R$ に線形分離可能
識別関数 $g_i(\mathbb{X}) = \mathbb{W}^{(i)}\cdot\mathbb{Y} \quad i=1,\cdots,R$
i category の $\mathbb{Y}$ について
$g_i(\mathbb{X}) < g_j(\mathbb{X})$ のとき
$$\begin{cases} \mathbb{W}^{(i)\prime} = \mathbb{W}^{(i)} + c\,\mathbb{Y} \\ \mathbb{W}^{(j)\prime} = \mathbb{W}^{(j)} - c\,\mathbb{Y} \end{cases}$$

4.6 applications to Φ machines

Φ machine の訓練にも用いられる

Chapter 5

5.1 the fundamental training theorem

— theorem 5・1

仮定 訓練部分集合 $\mathcal{Y}_1$ と $\mathcal{Y}_2$ が線形分離可能

S_wは、初期重みベクトル $\mathbb{W}_1$, 固定増分法、訓練列 S_y による重みベクトル列とする。

↓

有限 k_0 で

$$\mathbb{W}_{k_0} = \mathbb{W}_{k_0+1} = \mathbb{W}_{k_0+2} = \cdots\cdots$$

5.2 notation

$$\mathcal{Y}_2^* \rightarrow \mathcal{Y}_2' = \{ -Y_1^{(2)}, -Y_2^{(2)}, \cdots, -Y_{N_2}^{(2)} \}$$

調整された訓練集合 $\mathcal{Y}' = \mathcal{Y}_1 + \mathcal{Y}_2'$

$Y \cdot \mathbb{W} > 0$ 但し $Y \in \mathcal{Y}'$ なる $\mathbb{W}$ のあるとき、$\mathcal{Y}'$ は"線形に含まれている"

$S_{y'} = Y_1', Y_2', \cdots Y_k' \cdots\cdots$

$Y_k' \cdot \mathbb{W}_k > 0$ なる Y_k' を略すと

既約訓練列 $S_{\hat{y}} = \hat{Y}_1, \hat{Y}_2 \cdots\cdots$

$S_{\hat{w}} = \hat{\mathbb{W}}_1, \hat{\mathbb{W}}_2 \cdots\cdots$

$S_{\hat{w}}$ に繰返しがないから、

$\mathbb{W}_{k_0+1} = \mathbb{W}$ ならば 終結。

theorem 5.1 の proof $\Longrightarrow$ 終結のあることを proof

5.3 proof 1

$$\begin{cases} |\hat{\mathbb{W}}_{k+1}|^2 \geq \dfrac{(\hat{\mathbb{W}}_{k+1} \cdot \mathbb{W})^2}{|\mathbb{W}|^2} = \dfrac{1}{|\mathbb{W}|^2} \left(\sum_{j=1}^{k} \hat{Y}_j \cdot \mathbb{W} \right)^2 \geq \dfrac{(ka)^2}{|\mathbb{W}|^2} \\ |\hat{\mathbb{W}}_{k+1}|^2 \leq \sum_{j=1}^{k} |\hat{Y}_j|^2 \leq kM \end{cases}$$

5.4 proof 2

$\mathcal{W}$; 重みベクトルの解領域 ; 原点に頂点をもつ凸多面すい体

$\mathcal{W}' \subset \mathcal{W}$

$$\mathbb{W} \cdot \mathbb{Y} > \frac{M+b}{2} \quad ; b > 0, \; M = \max_{\mathbb{Y} \in \mathcal{Y}} |\mathbb{Y}|^2 \quad \text{—①}$$

$\mathbb{W} \in \mathcal{W}'$ として

$$|\mathbb{W} - \hat{\mathbb{W}}_j|^2 = \mathbb{W} \cdot \mathbb{W} - 2\mathbb{W} \cdot \mathbb{W}_j + \mathbb{W}_j \cdot \mathbb{W}_j$$

ここで

$$d_{k+1} = |\mathbb{W} - \hat{\mathbb{W}}_k|^2 - |\mathbb{W} - \hat{\mathbb{W}}_{k+1}|^2$$

$\rightarrow \hat{\mathbb{W}}_{k+1} = \hat{\mathbb{W}}_k + \hat{\mathbb{Y}}_k$ かつ, $\hat{\mathbb{W}}_k \cdot \hat{\mathbb{Y}}_k \leq 0$

$$\therefore d_{k+1} \geq 2\mathbb{W} \cdot \hat{\mathbb{Y}}_k - \hat{\mathbb{Y}}_k \cdot \hat{\mathbb{Y}}_k \geq 2\mathbb{W} \cdot \hat{\mathbb{Y}}_k - M \quad \text{—②}$$

①, ② より

$$d_{k+1} \geq b > 0$$

5.5

a training theorem for R-category linear machines

$\mathcal{Y}_1, \mathcal{Y}_2, \cdots \mathcal{Y}_R$

◦ 各 $\mathbb{Y} \in \mathcal{Y}_i$ に対し $\mathbb{Y} \cdot \mathbb{W}_i > \mathbb{Y} \cdot \mathbb{W}_j \quad (j = 1, \cdots R, \; j \neq i)$

ならば↓

$\{\mathcal{Y}_1, \mathcal{Y}_2, \cdots \mathcal{Y}_R\}$ は *linearly separable*

ならば↓

望ましい分類を実行する線形機械が存在する.

◦ 一般化された固定増分誤り訂正訓練手順

$S_{W_1} = \mathbb{W}_1^{(1)}, \mathbb{W}_2^{(2)}, \cdots$

$\vdots$

$S_{W_R} = \mathbb{W}_R^{(1)}, \mathbb{W}_R^{(2)}, \cdots$

(a) $\mathbb{W}_i^{(k)} \cdot \mathbb{Y}_k > \mathbb{W}_j^{(k)} \cdot \mathbb{Y}_k \quad \mathbb{Y}_k \in \mathcal{Y}_i, \; j = 1, \cdots R, \; j \neq i \Rightarrow \mathbb{W}_j^{(k+1)} = \mathbb{W}_j^{(k)}$

(b) $\mathbb{W}_l^{(k)} \cdot \mathbb{Y}_k \geq \mathbb{W}_j^{(k)} \cdot \mathbb{Y}_k$ 〃 $l, j = 1 \cdots R, \; j \neq l \neq i$ $\quad j = 1, \cdots R$

$$\Rightarrow \mathbb{W}_i^{(k+1)} = \mathbb{W}_i^{(k)} + c\,\mathbb{Y}_k$$

$$\mathbb{W}_l^{(k+1)} = \mathbb{W}_l^{(k)} - c\,\mathbb{Y}_k$$

$$\mathbb{W}_j^{(k+1)} = \mathbb{W}_j^{(k)}$$ 但, $j = 1 \cdots R, \; j \neq l, \; j \neq i$

(a) を生起するパターンを除去.

$S_{\hat{y}}$: 縮約訓練列

$S_{\hat{w}_j}$: 〃 重みベクトル列

定理 5.2

条件 $\mathcal{Y}=\{\mathcal{Y}_1,\cdots \mathcal{Y}_R\}$ は *linearly separable*

$S_{w_1},\cdots S_{w_R}$ は. 任意の初期重みベクトルで始まり. 一般化された固定増分誤り訂正手順を使って任意の訓練列 S_{y} による. 重みベクトル列とする.

このとき、~~ある~~最終的に重みベクトル列は. 解重みベクトル集合を導く.

→ *Proof*

step.1. $\mathcal{Z}$ をつくる

$\mathcal{Z}_{ij}$ は (y_j) $j=1\ldots R$, $j\neq i$, $y_j\in\mathcal{Y}_i$

D次元の R 個ブロックに分けて考える; RD次元を.

i番目の D成分のブロック $=y$

j 〃 $=-y$

その他の成分は zero

仮定 $\mathcal{Y}$ が $w_1,\cdots w_R$ で *linearly separable*

$\mathcal{Z}$ は RD次元ベクトル $W=(w_1,\cdots w_R)$ で *linearly contained*

この仮定の逆も真.

step 2. $S_{\mathcal{Z}}$ をつくる.

$\hat{y}_k\in\mathcal{Y}_i$ のとき. 次のような l が存在

$$\hat{w}_l^{(k)}\cdot\hat{y}_k \geq \hat{w}_j^{(k)}\hat{y}_k \qquad j=1\cdots R,\ j\neq l$$

このとき

$\mathcal{Z}_k=\mathcal{Z}_{il}(\hat{y}_k)$ とする。これにより

$S_{\hat{y}}\to S_z$ をつくる

step.3. S_w をつくる

$$W_k=[\hat{w}_1^{(k)},\cdots,\hat{w}_R^{(k)}]$$

5.6 $k=2$ の場合の関連訓練定理

定5.3 の proof ?