判別分析と特徴空間

潜在空間

本谷 秀堅

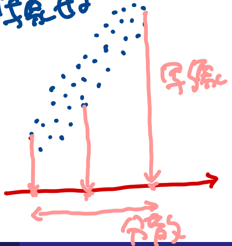
名古屋工業大学

線形
教師あり

復習: 主成分分析

問い：
分散が大きくなるように低次元に
写像せよ

・線形
・教師なし.



主成分分析

主成分分析

高次元データ分布を低次元空間に射影して近似する（教師なし）

主成分分析

高次元空間におけるデータの分布を最も良く近似する線形部分空間を求める。
近似の善し悪しを下記の基準で測る

- 分散最大基準
- 平均二乗誤差最小基準

準備

- D 次元データの集合 $\mathcal{D} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$
- データの平均 $\boldsymbol{m}$

$$\boldsymbol{m} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i$$

- $d\ (< D)$ 次元部分空間の正規直交基底, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d\}$
- d 次元部分空間への $D \times d$ 変換行列, $\mathbf{A}$

$$\mathbf{A} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d].$$

  - 正規直交基底なので次式が成立する

$$\boldsymbol{v}_i^\top \boldsymbol{v}_j = \begin{cases} 1, & (i = j) \\ 0, & (i \neq j) \end{cases}, \quad \mathbf{A}^\top \mathbf{A} = \mathbf{I}.$$

- D 次元データ $\boldsymbol{x}_i$ を d 次元ベクトル $\boldsymbol{y}_i$ へと変換:

$$\boldsymbol{y}_i = \mathbf{A}^\top \boldsymbol{x}_i$$

行列 A による射影（手書き）

分散最大基準: 準備

- d 次元ベクトルへの変換後のデータ集合, $\mathcal{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\}$
- $\mathcal{Y}$ の平均, $\tilde{\boldsymbol{m}}$

$$\tilde{\boldsymbol{m}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i = \mathbf{A}^\top \boldsymbol{m}$$

- 射影後の分散, $\tilde{\sigma}^2(\mathbf{A})$, は次式のとおり：

$$\tilde{\sigma}^2(\mathbf{A}) = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{y}_i - \tilde{\boldsymbol{m}})^\top (\boldsymbol{y}_i - \tilde{\boldsymbol{m}}) = \mathrm{tr}(\mathbf{A}^\top \boldsymbol{\Sigma} \mathbf{A})$$

ただし, $\boldsymbol{\Sigma}$ は $\mathcal{D}$ の共分散行列

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^\top$$

前項の導出

$$\frac{1}{N}\sum_i (\boldsymbol{y}_i - \tilde{\boldsymbol{m}})^T (\boldsymbol{y}_i - \tilde{\boldsymbol{m}}) =$$

$$= \frac{1}{N}\sum_i \operatorname{tr}\left(\boldsymbol{A}^T(\boldsymbol{x} - \boldsymbol{m})(\boldsymbol{A}^T(\boldsymbol{x} - \boldsymbol{m}))^T\right)$$

$$=$$

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \operatorname{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T), \quad \frac{\partial}{\partial \boldsymbol{A}}\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \boldsymbol{B}^T$$

分散最大基準による主成分分析の導出

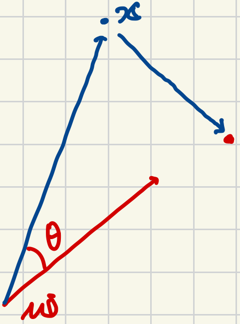
$\mathbf{A}^\top\mathbf{A} = \mathbf{I}$ を満たしつつ、分散 $\tilde{\sigma}(\mathbf{A})$ を最大にする $\mathbf{A}$ を求めたい

$$J(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^\top\boldsymbol{\Sigma}\mathbf{A}) - \mathrm{tr}((\mathbf{A}^\top\mathbf{A} - I)\mathbf{L})$$

ただし $\mathbf{L}$ は $d \times d$ の対角行列。上式を最大にする $\mathbf{A}^*$ は次の通りで、$\tilde{\sigma}^2$ の最大値は $\lambda_1 + \lambda_2 + \cdots + \lambda_d$ になる。

$$\mathbf{A}^* = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d], \quad \mathbf{L}^* = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{bmatrix}$$

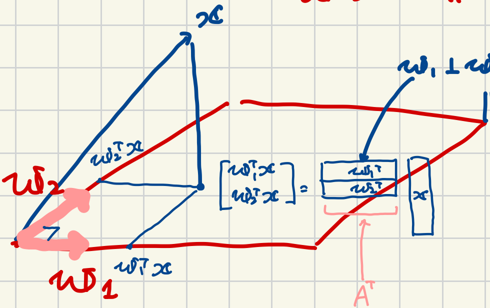
すなわち、共分散行列 $\boldsymbol{\Sigma}$ の固有値を大きい順に d 個選び、対応する d 個の固有ベクトルが張る空間が投影後の分散を最大にする。



ここを求める.

$\|w\| = 1$ とする

$w^T x = \|x\| \cos\theta$



$\left.\begin{array}{l} w_1 \perp w_2 \\ \|w\| = 1 \end{array}\right\} A^T A = I$

$$\begin{bmatrix} w_1^T x \\ w_2^T x \end{bmatrix} = \underbrace{\begin{bmatrix} w_1^T \\ w_2^T \end{bmatrix}}_{A^T} \begin{bmatrix} x \end{bmatrix}$$

Aの階数

射影後の分散: $\mathrm{trace}(A^T \Sigma A)$

$$= \mathrm{trace}\left(\boxed{A^T}\ \boxed{\Sigma}\ \boxed{A}\right)$$

$$\boxed{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)(x_i - m)^T$$

データ空間での
共分散

前項の導出

$$J(\mathbf{A}) = \underbrace{\mathrm{tr}(\mathbf{A}^\top \boldsymbol{\Sigma} \mathbf{A})}_{\text{射影後の分散}} - \underbrace{\mathrm{tr}((\mathbf{A}^\top \mathbf{A} - I)\mathbf{L})}_{A^TA=I\text{の拘束}}$$

$L = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{bmatrix}$ ラグランジュ係数

$$\frac{\partial J(\boldsymbol{A})}{\partial \boldsymbol{A}} = 2\Sigma A - 2AL = \mathbf{0}$$

ただし下記を利用する

$\Sigma A = AL$

$$\frac{\partial}{\partial \boldsymbol{X}} \mathrm{tr}(\boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{X}) = (\boldsymbol{Y} + \boldsymbol{Y}^T)\boldsymbol{X}$$

$A = [\, w_1 \mid w_2 \,]$

上記より次式が成り立つ

$$\boldsymbol{A}^T \boldsymbol{\Sigma} \boldsymbol{A} = \boldsymbol{L}$$

$\Sigma w = \lambda w$

従って

$\Sigma A = AL$

左から A^T

$$\max_{\boldsymbol{A}} \underbrace{\tilde{\sigma}^2(\boldsymbol{A})}_{\text{射影後の分散}} = \max_{\boldsymbol{A}} \left\{ \underbrace{\mathrm{tr}\left(\boldsymbol{A}^T \boldsymbol{\Sigma} \boldsymbol{A}\right)}\right\} =$$

Σの固有ベクトルが $\mathrm{tr}(A^T\Sigma A)$を極大にする。

$$= \max_{A} \{ \mathrm{trace}(L) \}$$

$$= \max \{ \underbrace{\lambda_1 + \lambda_2 + \cdots + \lambda_d}_{\text{固有値を大きい順に}d\text{個選ぶ.}} \}$$

写影先の次元
(我々が勝手に決める)

平均二乗誤差最小基準：準備

- データ $\boldsymbol{x}_i$ $(i = 1, 2, \ldots, N)$ を、点 $\boldsymbol{x}_0$ を通る d 次元部分空間に正射影した後の座標は次式のとおり：

$$\mathbf{A}\boldsymbol{y}_i + \boldsymbol{x}_0 = \mathbf{A}\mathbf{A}^\top(\boldsymbol{x}_i - \boldsymbol{x}_0) + \boldsymbol{x}_0$$

- 射影前後の差を表すベクトルは次のとおり

$$(\mathbf{A}\boldsymbol{y}_i + \boldsymbol{x}_0) - \boldsymbol{x}_i = (\mathbf{A}\mathbf{A}^\top - \mathbf{I})(\boldsymbol{x}_i - \boldsymbol{x}_0)$$

- $\mathbf{Q} = \mathbf{I} - \mathbf{A}\mathbf{A}^\top$ と置く。平均二乗誤差, ϵ^2, は次のとおり

$$\epsilon^2(\mathbf{A}, \boldsymbol{x}_0) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{x}_0)^\top\mathbf{Q}(\boldsymbol{x}_i - \boldsymbol{x}_0)$$

演習

ϵ^2 の式を導出せよ。

平均二乗誤差最小基準

二乗誤差を最小にするためには、点 $\boldsymbol{x}_0$ は次式を満たさなければいけない。

$$\frac{\partial \epsilon^2}{\partial \boldsymbol{x}_0} = \frac{1}{N}\sum_{i=1}^{N}(2\mathbf{Q}\boldsymbol{x}_0 - 2\mathbf{Q}\boldsymbol{x}_i) = \mathbf{0}$$

よって

$$\mathbf{Q}\boldsymbol{x}_0 = \mathbf{Q}\boldsymbol{m}.$$

このとき次式が成立する（分散最大化の問題と同じ問題に帰着される）

$$\epsilon^2(\mathbf{A}) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{m})^\top \mathbf{Q}(\boldsymbol{x}_i - \boldsymbol{m}) = \mathrm{tr}\boldsymbol{\Sigma} - \mathrm{tr}(\mathbf{A}^\top \boldsymbol{\Sigma}\mathbf{A})$$

前項の導出

$$
\begin{aligned}
\epsilon^2(\boldsymbol{A}) &= \frac{1}{N}\sum_i (\boldsymbol{Q}(\boldsymbol{x}_i - \boldsymbol{m}))^T(\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{m})) \\
&= \frac{1}{N}\sum_i (\boldsymbol{x}_i - \boldsymbol{m})^T\boldsymbol{Q}(\boldsymbol{x}_i - \boldsymbol{m}) \\
&= \frac{1}{N}\sum_i (\boldsymbol{x}_i - \boldsymbol{m})^T(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T)(\boldsymbol{x}_i - \boldsymbol{m}) \\
&= \frac{1}{N}\sum_i \mathrm{tr}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^T - \mathrm{tr}(\boldsymbol{A}^T(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^T\boldsymbol{A}) \\
&= \mathrm{tr}(\boldsymbol{\Sigma}) - \mathrm{tr}(\boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A}) \qquad (1)
\end{aligned}
$$

平均二乗誤差最小基準

$\mathbf{A}^\top\mathbf{A} = \mathbf{I}$ を満たしつつ平均二乗誤差を最小にするには、$\mathrm{tr}(\mathbf{A}^\top\mathbf{\Sigma}\mathbf{A})$ を最大にすれば良い。分散最大基準と同じ。

すなわち、共分散行列 $\mathbf{\Sigma}$ の固有値を大きい順に d 個選び、対応する d 個の固有ベクトルが張る部分空間に正射影するとき、平均二乗誤差が最小になる。

· プリント採点.
· Moodle にPDF upload

→ 4/17

判別分析

教師あり学習

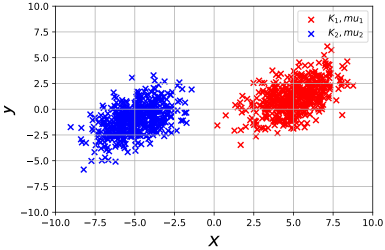
◎ 写像 (mapping)
◎ 特徴抽出
◎ 潜在表現 (latent representation)

Fisher 判別分析

Fisher 判別分析

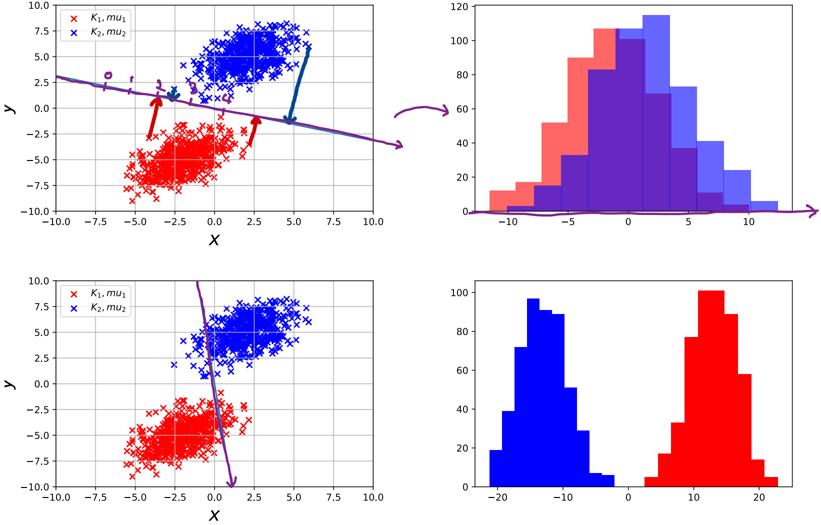
高次元データ分布を識別に適した低次元空間に射影して近似する（教師あり）

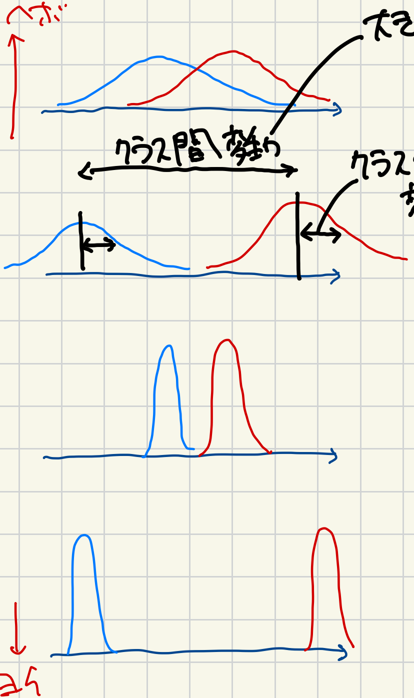
例



どの向きに射影すれば赤と青を分離できるか？

例





大きい程良い

小さい程良い

線形分離可能性

$$\frac{\text{クラス間変動}}{\text{クラス内変動}} \rightarrow \max$$

Fisher 判別分析：準備

クラス内変動とクラス間変動の比を最大にする 1 次元空間を求める

- クラス ω_1 と ω_2 に属する学習データの集合を $\mathcal{X}_1$ と $\mathcal{X}_2$ で表す。
- それぞれの学習データ数を $|\mathcal{X}_1| = N_1$, $|\mathcal{X}_2| = N_2$ で表す。
- 全体の平均を $\boldsymbol{m}$ で表す ($N = N_1 + N_2$):

$$\boldsymbol{m} = \frac{1}{N}\sum \boldsymbol{x}$$

- それぞれの学習データの平均を $\boldsymbol{m}_i$ で表す:

$$\boldsymbol{m}_1 = \frac{1}{N_1}\sum_{\boldsymbol{x}\in\mathcal{X}_1} \boldsymbol{x}, \quad \boldsymbol{m}_2 = \frac{1}{N_2}\sum_{\boldsymbol{x}\in\mathcal{X}_2} \boldsymbol{x}$$

- クラス ω_i の変動を表す変動行列を $\mathbf{S}_i$ で表す:

$$\mathbf{S}_i = \sum_{\boldsymbol{x}\in\mathcal{X}_i} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^\top$$

クラス i　クラス i の平均

Fisher 判別分析：準備

データから直接計算可 ⇒ 射影後の分散をここから計算したい。

- クラス内変動行列 (within-class scatter matrix), $\mathbf{S}_{\mathrm{W}}$

$$\boxed{\mathbf{S}_{\mathrm{W}}} = \underline{\mathbf{S}_1} + \underline{\mathbf{S}_2} = \sum_{i=1}^{2}\sum_{\boldsymbol{x}\in\mathcal{X}_i}(\boldsymbol{x}-\boldsymbol{m}_i)(\boldsymbol{x}-\boldsymbol{m}_i)^\top$$

Sw

- クラス間変動行列 (between-class scatter matrix), $\mathbf{S}_{\mathrm{B}}$

$$\boxed{\mathbf{S}_{\mathrm{B}}} = \sum_{i=1}^{2} N_i(\underline{\boldsymbol{m}_i} - \underline{\boldsymbol{m}})(\boldsymbol{m}_i - \boldsymbol{m})^\top = \frac{N_1 N_2}{N}(\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^\top$$

クラスiの平均　全体の平均

写像して1次元に

- $D \times 1$ 変換ベクトル, $\boldsymbol{w}$

$$y = \boldsymbol{w}^\top \boldsymbol{x}$$

y = w^T x

- $\boldsymbol{x} \in \mathcal{X}_i$ を変換して得られた y の集合 $\mathcal{Y}_i$

クラス内・外の変動に興味

Fisher 判別分析：準備

クラスiの写影後の平均は

平均を写影

- 変換後の平均, $\tilde{m}_i$:

$$\tilde{m}_i = \frac{1}{N_i}\sum_{y\in\mathcal{Y}_i} y = \frac{1}{N_i}\sum_{\boldsymbol{x}\in\mathcal{X}_i} \boldsymbol{w}^\top \boldsymbol{x} = \boldsymbol{w}^\top \boldsymbol{m}_i$$

知りたい

- 変換後のクラス内変動 $\tilde{S}_\mathrm{W}$

写影後の w を変化させて

小さくしたい

$$\tilde{S}_\mathrm{W} = \tilde{S}_1 + \tilde{S}_2 = \sum_{i=1}^{2} \underbrace{\sum_{y\in\mathcal{Y}_i} (y-\tilde{m}_i)^2}_{} = \boldsymbol{w}^\top \mathbf{S}_\mathrm{W} \boldsymbol{w}$$

データから得られる

クラス内の写影後の変動

- 変換後のクラス間変動 $\tilde{S}_\mathrm{B}$

$$\tilde{S}_\mathrm{B} = \sum_{i=1}^{2} N_i(\tilde{m}_i - \tilde{m})^2 = \frac{N_1 N_2}{N}(\tilde{m}_1 - \tilde{m}_2)^2 = \boldsymbol{w}^\top \mathbf{S}_\mathrm{B} \boldsymbol{w}$$

大きくしたい

Fisher 判別分析

分子分母に
スケール不定性

- Fisher 判別基準（クラス内変動・クラス間変動比） $J(\boldsymbol{w})$:

$$\boxed{J(\boldsymbol{w}) = \frac{\tilde{S}_{\mathrm{B}}}{\tilde{S}_{\mathrm{W}}} = \frac{\boldsymbol{w}^{\top}\mathbf{S}_{\mathrm{B}}\boldsymbol{w}}{\boldsymbol{w}^{\top}\mathbf{S}_{\mathrm{W}}\boldsymbol{w}}}$$

を最大に

Fisher 判別基準を最大にする $\boldsymbol{w}$ は次式で与えられる

$$\boldsymbol{w} \propto \mathbf{S}_{\mathrm{W}}^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2)$$

$$\underset{w}{[\;]} = \boxed{S_W^{-1}} \left(\underset{m_1}{[\;]} - \underset{m_2}{[\;]} \right)$$

Fisher 判別基準を最大にするベクトル $\boldsymbol{w}$ の式を導出せよ。

- スケール不定性があるので $w^T S_W w = 1$ の拘束下で $w^T S_B w$ を最大にする。

↑楽になる。

- $J(w) = w^T S_B w - \lambda (w^T S_W w - 1)$ を maxに。

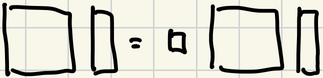
$$\boxed{\frac{1}{2}\frac{dJ}{dw} = S_B w - \lambda S_W w = 0}$$

$\boldsymbol{S}_{\mathrm{W}}^{-1}\boldsymbol{S}_{\mathrm{B}}$ の最大固有値を λ_1 とおくと $J(\boldsymbol{w})$ の最大値は λ_1 になる。

内積 $(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T\boldsymbol{w}$ がスカラであることに注意すると次式を得ることが出来る

$$\boldsymbol{w} \propto \boldsymbol{S}_{\mathrm{W}}^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2)$$

$$S_B w = \lambda S_W w$$



$$S_W^{-1} S_B w = \lambda w$$

必要条件

最大固有値に対応する

最適な w は $S_W^{-1} S_B$ の固有ベクトル

$$\frac{w^T S_B w}{w^T S_W w} = \frac{w \lambda S_W w}{w^T S_W w} = \lambda$$

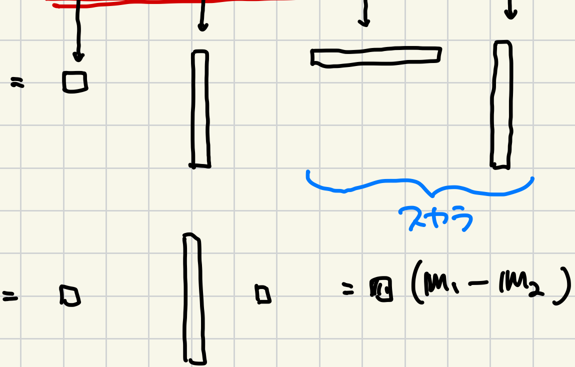
最大固有値を我々は選ぶべき.

$$\underline{\underline{S_B}}\, w = \lambda S_W w$$

$$S_B = \frac{N_1 N_2}{N} (m_1 - m_2)(m_1 - m_2)^T$$

?

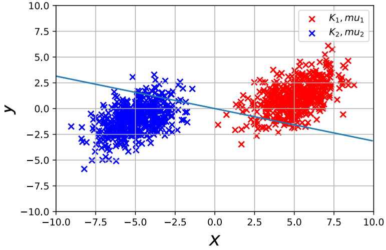
$$\lambda S_W w = \frac{N_1 N_2}{N} (m_1 - m_2)(m_1 - m_2)^T w$$

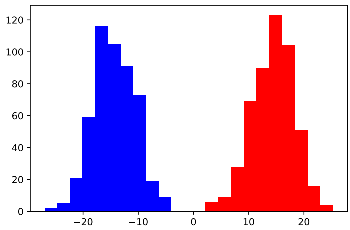


$$\underset{\lambda}{\square}\, \boxed{S_w}\, \boxed{w} = \square\, \underset{m_1 - m_2}{\boxed{}}$$

$$\boxed{\boxed{w} \propto S_w^{-1}(m_1 - m_2)}$$

例題の解答





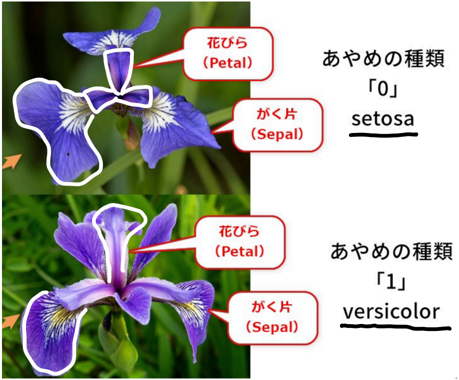
$$\boldsymbol{m}_1 = \begin{bmatrix} 5.04 \\ 1.12 \end{bmatrix}, \boldsymbol{S}_1 \propto \begin{bmatrix} 2.03 & 1.10 \\ 1.10 & 2.22 \end{bmatrix}, \quad \boldsymbol{m}_2 = \begin{bmatrix} -5.00 \\ -1.00 \end{bmatrix}, \boldsymbol{S}_2 \propto \begin{bmatrix} 1.88 & 0.93 \\ 0.93 & 2.04 \end{bmatrix}$$

$$\boldsymbol{w} \propto \boldsymbol{S}_{\mathrm{w}}^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2) = \begin{bmatrix} 3.07 \\ -0.97 \end{bmatrix}$$

識別器と識別に有利な特徴

準備：データの分布

例：あやめの種類を花の形で識別したい



https://atmarkit.itmedia.co.jp/ait/articles/2206/13/news032.html

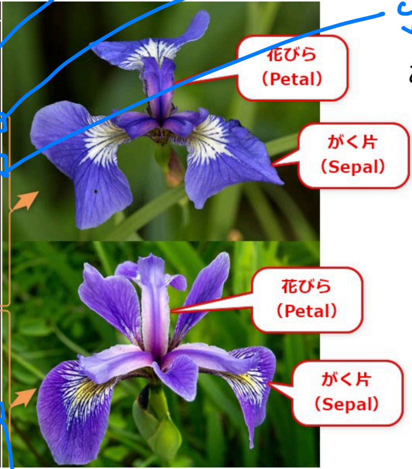
例：あやめのデータ

$x_0 \cdots x_2 \cdots$　$y_0 = 0$　$y_2 = 0$

列番号→	1	2	3	4	目的変数↓
説明変数→	Sepal Length	Sepal Width	Petal Length	Petal Width	target
行番号↓	がく片の長さ (cm)	がく片の幅 (cm)	花びらの長さ (cm)	花びらの幅 (cm)	あやめの種類
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
5	5.4	3.9	1.7	0.4	0
6	4.6	3.4	1.4	0.3	0
7	5.0	3.4	1.5	0.2	0
8	4.4	2.9	1.4	0.2	0
49	5.0	3.3	1.4	0.2	0
50	7.0	3.2	4.7	1.4	1
51	6.4	3.2	4.5	1.5	1
52	6.9	3.1	4.9	1.5	1
53	5.5	2.3	4.0	1.3	1
54	6.5	2.8	4.6	1.5	1
55	5.7	2.8	4.5	1.3	1
56	6.3	3.3	4.7	1.6	1
57	4.9	2.4	3.3	1.0	1

花の数

X　$y_{55} = 1$



あやめの種類
「0」
setosa

あやめの種類
「1」
versicolor

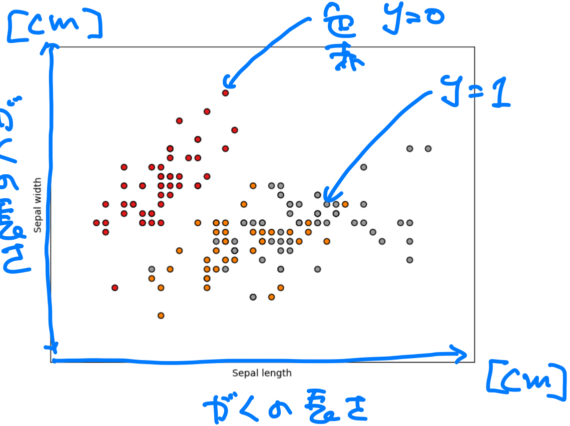
https://atmarkit.itmedia.co.jp/ait/articles/2206/13/news032.html

準備：データの分布

がく
片

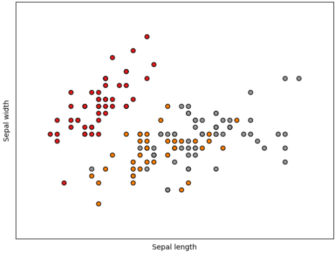
説明変数→ 行番号↓	Sepal Length がく片の長さ (cm)	Sepal Width がく片の幅 (cm)	Petal Length 花びらの長さ (cm)	Petal Width 花びらの幅 (cm)	target あやめの種類
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
5	5.4	3.9	1.7	0.4	0
6	4.6	3.4	1.4	0.3	0
7	5.0	3.4	1.5	0.2	0
8	4.4	2.9	1.4	0.2	0
49	5.0	3.3	1.4	0.2	0
50	7.0	3.2	4.7	1.4	1
51	6.4	3.2	4.5	1.5	1
52	6.9	3.1	4.9	1.5	1
53	5.5	2.3	4.0	1.3	1
54	6.5	2.8	4.6	1.5	1
55	5.7	2.8	4.5	1.3	1
56	6.3	3.3	4.7	1.6	1
57	4.9	2.4	3.3	1.0	1

データの分布（教師あり）

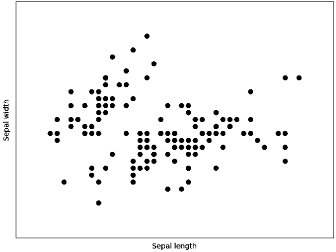


準備：データの分布

データの分布（教師あり）



データの分布（教師なし）



(最尤推定)

/ パラメータ
(これから学習で決める)

準備：線形識別器の教師あり学習

準備：ニューラルネットワークの「ユニット」

ニューロンのひとつひとつ

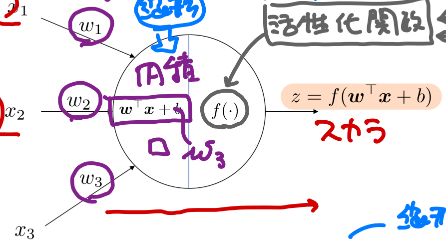
unit

活性化関数 ← 手作業で与える.

おくの長さ

入力　幅

1



線形

内積

w_3

出力

出力

シグモイド関数をつかうと

線形識別

ロジスティック回帰ができる.

ユニットの入出力

- 入力：$\boldsymbol{x} = (x_1, x_2, \ldots,)^\top$
- 出力：$y = f(u)$
  - まず重みとの内積＋定数を計算：
    $u = w_1 x_1 + w_2 x_2 + \cdots + b = \boldsymbol{w}^\top \boldsymbol{x} + b$
  - 次に活性化関数：$f(u) = f(\boldsymbol{w}^\top \boldsymbol{x} + b)$

$$P(y=1|x) = \frac{1}{1+\exp(-a)}$$

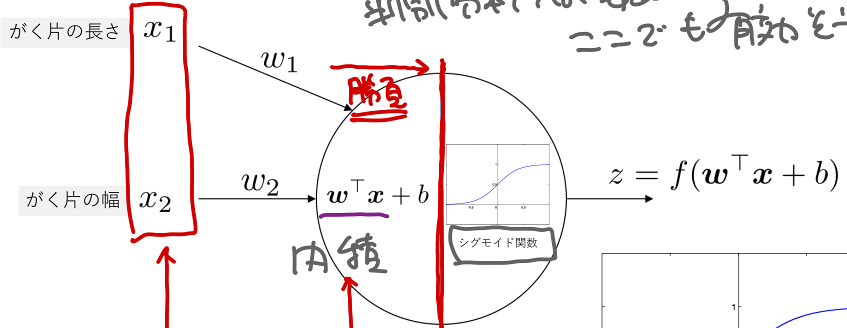
$$a = w^T x + w_0$$

ロジスティック回帰

準備：線形識別器の教師あり学習

例：あやめの識別タスク

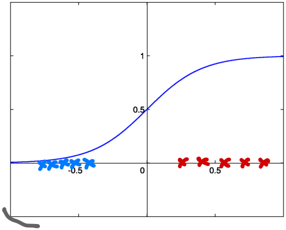
判別分析への理解は
ここでも有効そう.



閾値

内積

高次元



wが張る直線への写影として.

準備：線形識別器の教師あり学習

ω は 損失が最小になるように

(負の対数尤度由来.)

例：あやめの識別器の学習（2 クラス）

損失関数（交差エントロピー）：

$P(y|x) = \pi^y (1-\pi)^{1-y}$

$\log P = y \log \pi + (1-y) \log(1-\pi)$

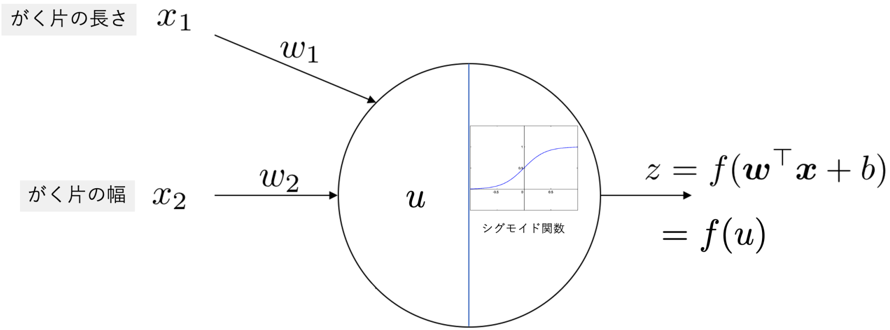
$$\sum_{i=1}^{N} y_i \log f(\omega_1 x_{i,1} + \omega_2 x_{i,2} + \omega_0) + (1 - y_i) \log(1 - f(\omega_1 x_{i,1} + \omega_2 x_{i,2} + \omega_0))$$

- $x_{i,1}$: i 番目のデータの「がく片の長さ」
- $x_{i,2}$: i 番目のデータの「がく片の幅」
- y_i : i 番目のデータの「あやめの種類」(0 or 1)
- $f(\cdot)$: データ $\boldsymbol{x}$ が「アヤメ 1」のものである事後確率の回帰

シグモイド関数

準備：線形識別器の教師あり学習

例：あやめの識別タスク（再掲）

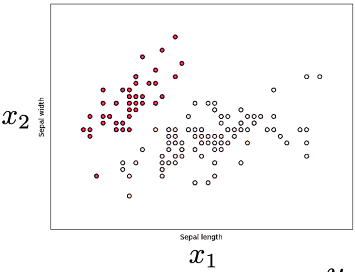


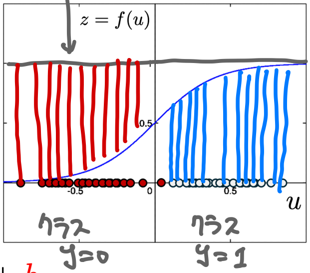
$$u = {\color{red}w_1} x_1 + {\color{red}w_2} x_2 + {\color{red}b}$$

準備：線形識別器の教師あり学習

w_0, w_1, w_2 を変化させて 青線 赤線 の長さを 最大に.

(1) 重み w_1, w_2 との積和でスカラ u に、(2) シグモイド関数で z に



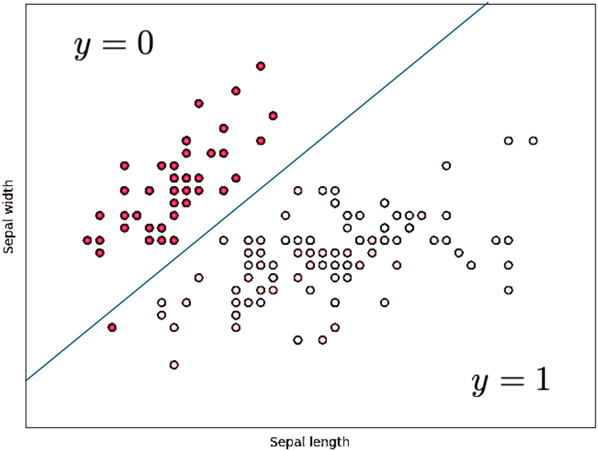


$$u = w_1 x_1 + w_2 x_2 + b$$

4/24

準備：線形識別器の教師あり学習

$z \lessgtr 0.5$ で識別



準備：線形識別器の教師あり学習

$z = f(u) \lessgtr 0.5$ で識別

識別関数

事後確率

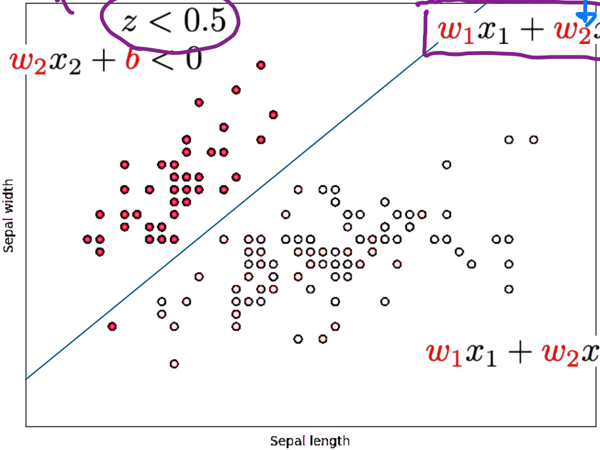
負

$z = 0.5$

$w_1 x_1 + w_2 x_2 + b = 0$

$z < 0.5$

$w_1 x_1 + w_2 x_2 + b < 0$



正

事後確率

$z > 0.5$

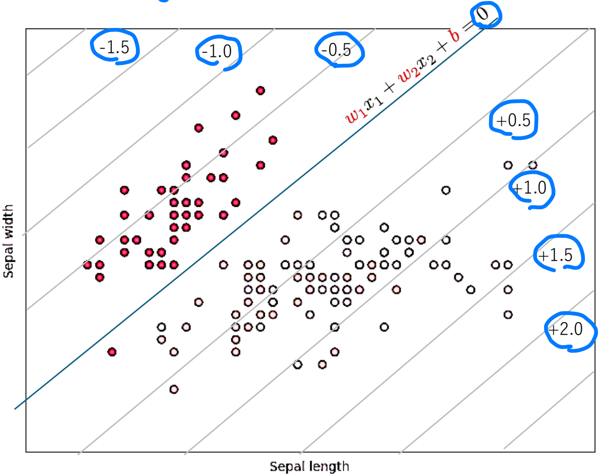
$w_1 x_1 + w_2 x_2 + b > 0$

$u = \boldsymbol{w}^\top x + b = 0$ が識別境界

準備：特徴空間への写像

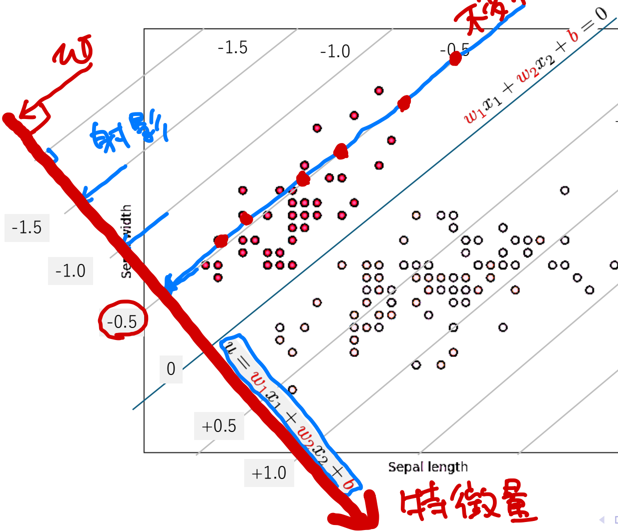
$u = \boldsymbol{w}^\top \boldsymbol{x} + b$ のデータ $\boldsymbol{x}$ から特徴 u への**写像**

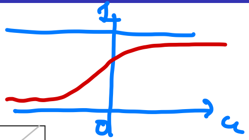
識別関数の等高線



準備：特徴空間への写像

$u = \boldsymbol{w}^\top \boldsymbol{x} + b$ のデータ $\boldsymbol{x}$ から特徴 u への写像

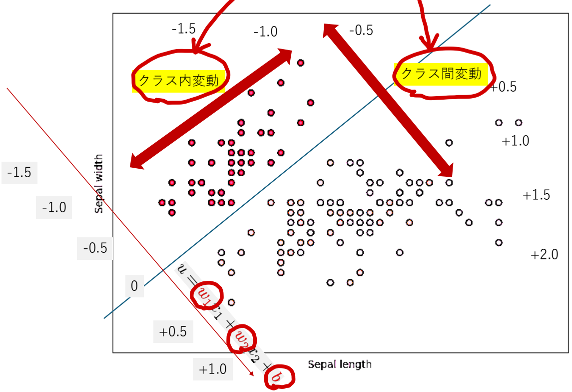




準備：特徴空間への写像

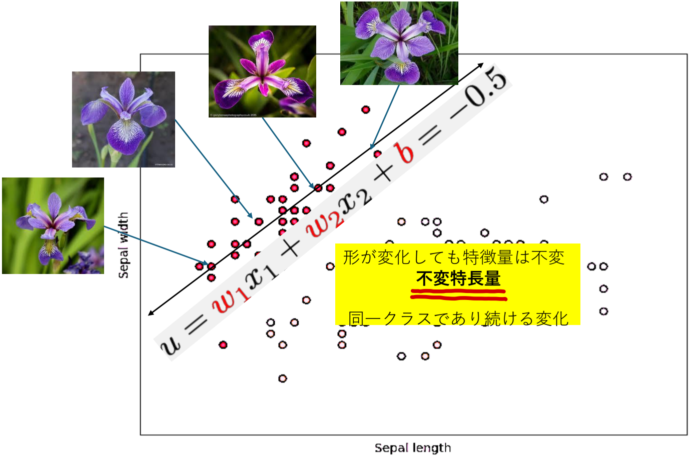
識別タスクに望ましい特徴量（参考：判別分析）

- クラス内変動に不変・鈍感
- クラス間変動に敏感



準備：特徴空間への写像

同一カテゴリであり続ける変換に対して鈍感・不変な特徴量が良い



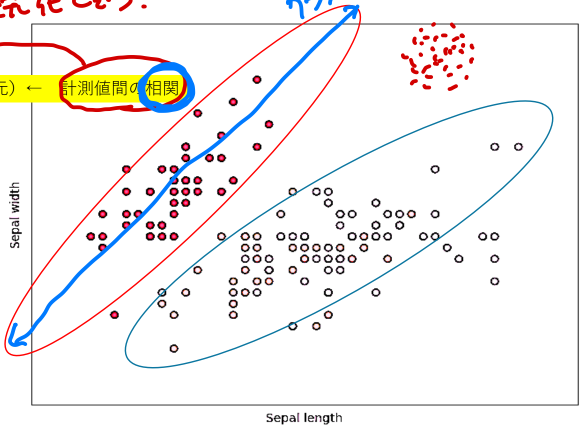
準備：特徴空間への写像

計測値間に相関により同一カテゴリの分布が低次元に

低次元化しよう.

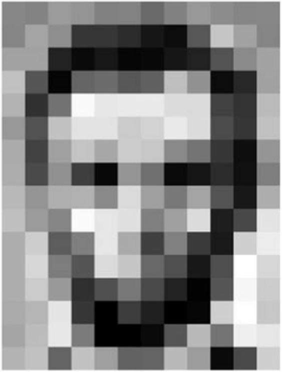
クラス内変動

分布が扁平（低次元）← 計測値間の相関



画像データ

- 画素値の配列（高次元データ）
- 近傍の画素値間に強い関係（高次元空間中で低次元に分布）

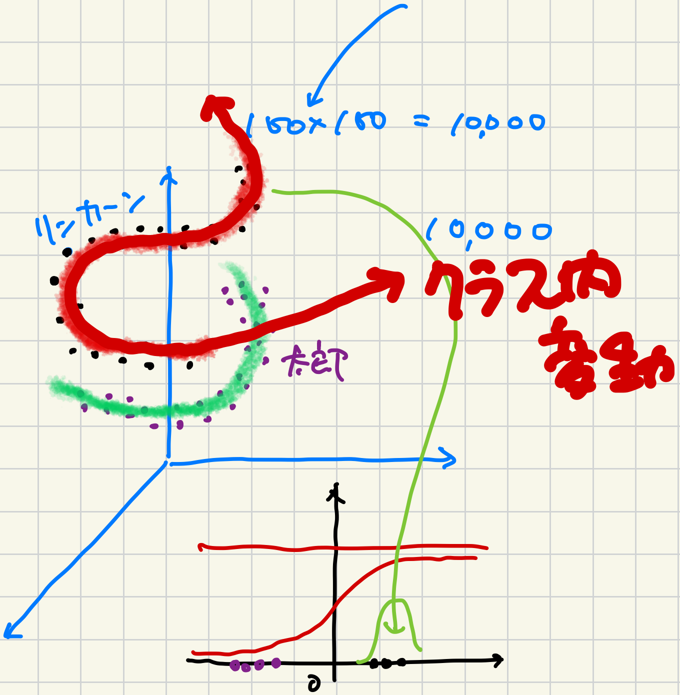


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

100

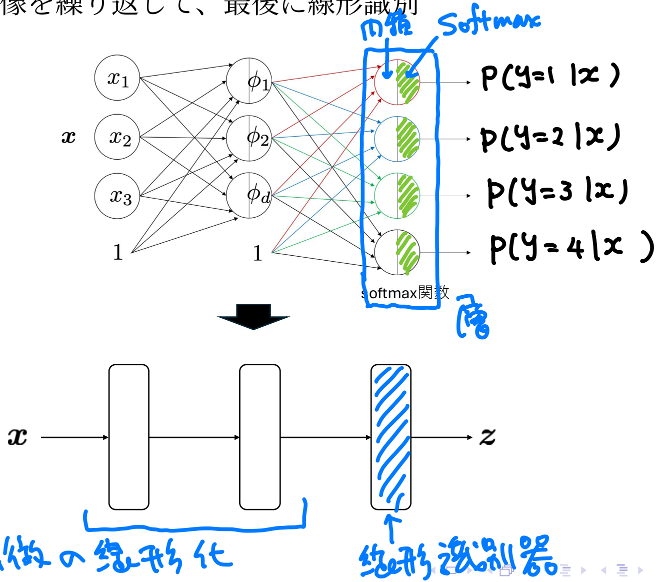
150

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218


100×100 = 10,000
リンカーン
10,000
クラス内
変動
0

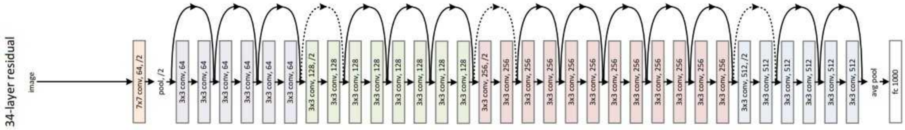
準備：ニューラルネットワーク

シンプルな写像を繰り返して、最後に線形識別



準備：ニューラルネットワーク

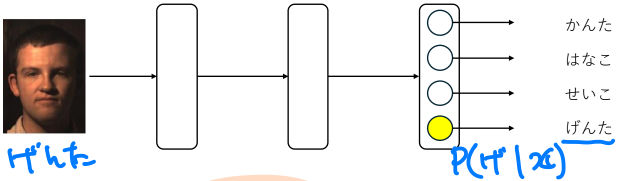
例：ResNet34 (ImageNet で学習。1000 クラス識別)

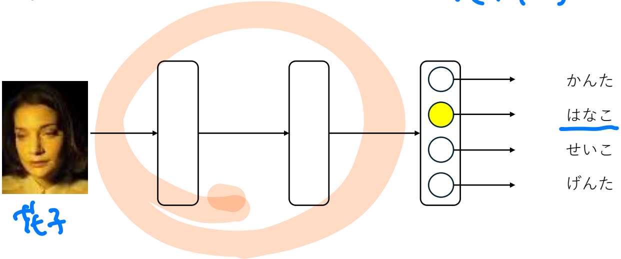


シンプルな写像を繰り返して、最後に線形識別

準備：識別に適した画像特徴量

多クラス識別：最終層のユニットのひとつが「発火」

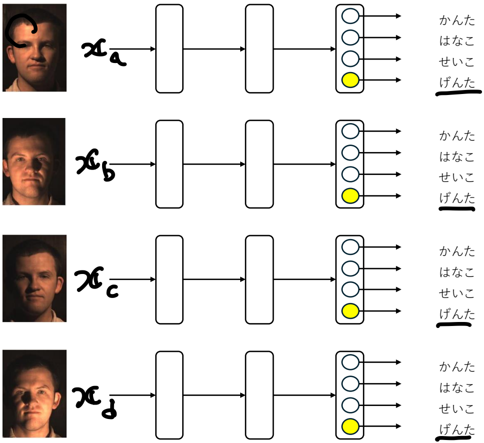




正解クラスの事後確率が最大

準備：識別に適した画像特徴量

見え方が変化しても同じユニットが発火



「おばあさん細胞」の発想

寄り道：ニューラルネットワークの発想

ウィキペディア
フリー百科事典

Wikipedia内を検索 検索

寄付

おばあさん細胞

文A 5の言語版

目次 非表示

ページ ノート　閲覧 編集 履歴を表示 ツール

出典: フリー百科事典『ウィキペディア（Wikipedia）』

おばあさん細胞（おばあさんさいぼう、英: Grandmother cell、またはgnostic neuron）とは、その人のおばあさん、より一般的に言えば、複合的な特定の概念や対象物を表現している仮想的な細胞である[1]。おばあさん細胞は、その人が自身のおばあさんの姿を見たり、声を聞くなどして感覚的に識別する際に活動する[2]。おばあさん細胞という用語はジェローム・レトビンによる造語である。

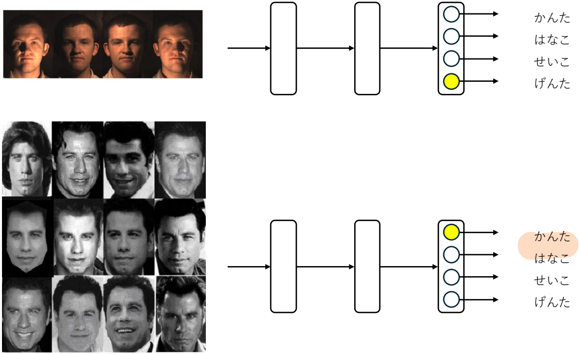
概要 [編集]

おばあさん細胞仮説を支持する初期の研究として、サルの下側頭皮質にある、手と顔に選択的に発火する視覚ニューロンに関する研究がある[3][4][5][6]。しかし、果物や性器などのサルにとって重要な他の視覚対象に対して選択的に発火する細胞は見つかっていない。このことは、バナナなどの他のカテゴリーを識別するよりも、顔を識別する方がサルにとってより重要であるためと考えられている。加えて、サルが識別しなくてはならない他の視覚刺激に比べ、顔は全体的な特徴と細部が顔同士でよく似ているためと考えられている[1]。

おばあさん細胞仮説を支持する最近の研究[7][8]として、下側頭皮質の細胞は任意の視覚対象に対して非常に選択的に反応するように訓練できるということを示した研究が存在する。この性質はおばあさん細胞に求められる条件に合致したものである。加えて、ヒトの海馬にある細胞が個人の顔[9]を含む、認識のカテゴリーに対して非常に選択的に反応する[10][11]といった証拠が見つかっている。

準備：教師あり学習

正解ユニットが発火するように学習



同じカテゴリであり続ける変化に鈍感・不変

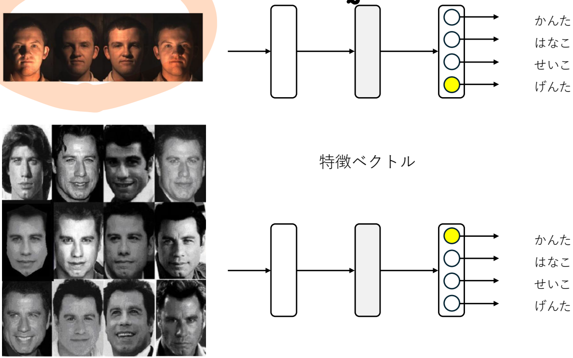
準備：教師あり学習

正解ユニットが発火するように学習

クラス内変動

できるだけ不変

鈍感に



同じカテゴリであり続ける変化に鈍感・不変な**特徴量の獲得**