

線形判別分析

Linear Discriminant Analysis

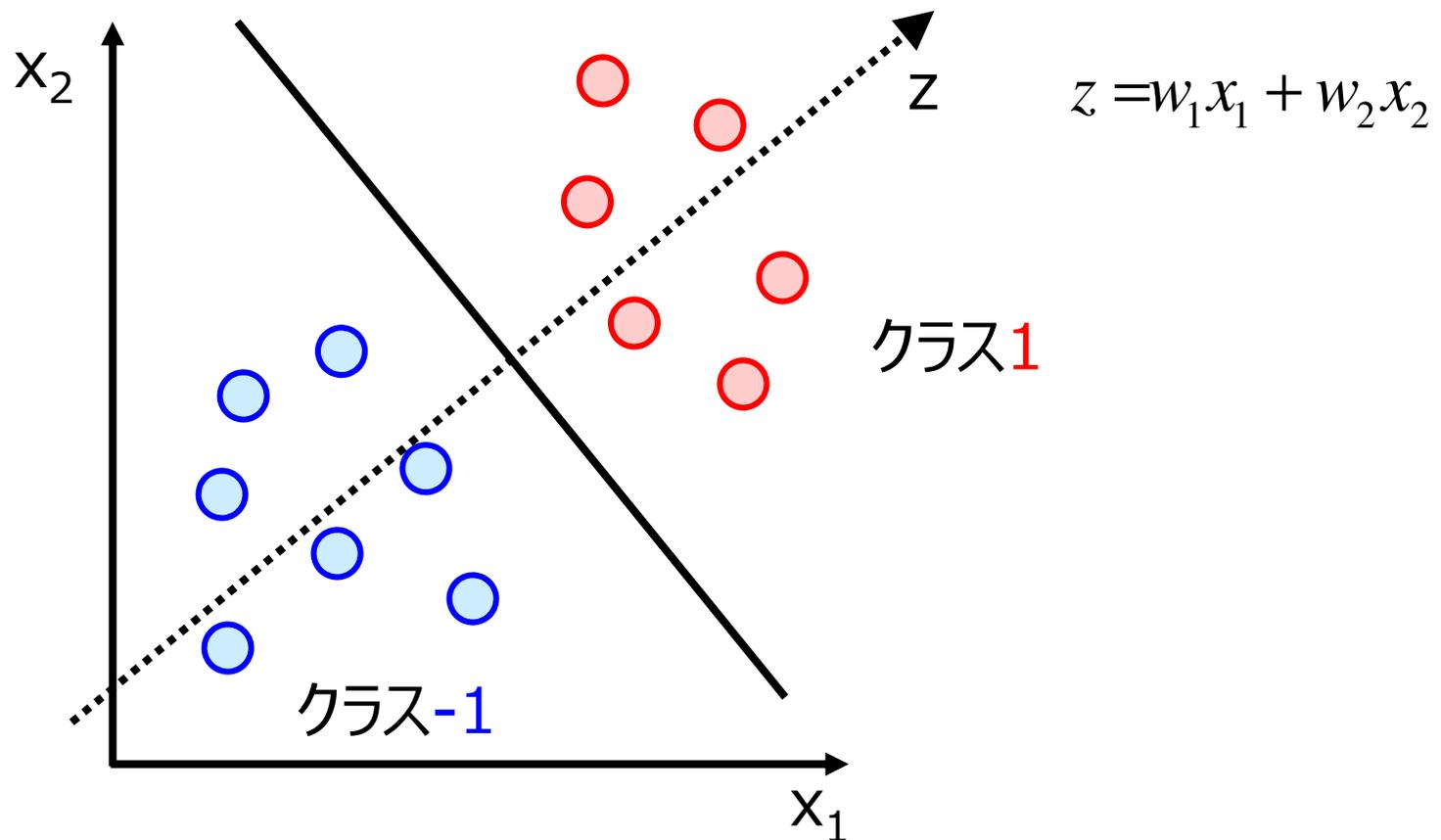
LDA

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

線形判別分析 (LDA) とは？

✓ 線形判別分析 (Linear Discriminant Analysis, LDA)

- 1次元(z)に線形写像し、 z で 2 つのクラスを識別する
- 2つのクラスを “最もよく判別する” ように線形写像する
- クラスが3つ以上あるときにも拡張できる



“最もよく判別する”とは？

- ✓① 各クラスのサンプルは固まっている

z でのクラス内のばらつき V_{Wz}

$$V_{Wz} = \sum_{i \in \text{クラス1}} (z^{(i)} - \bar{z}_{[1]})^2 + \sum_{i \in \text{クラス-1}} (z^{(i)} - \bar{z}_{[-1]})^2$$

$\bar{z}_{[k]}$: クラス k のみの z の平均

- ✓② クラス1(赤)とクラス-1(青)は散らばっている

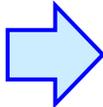
z でのクラス間のばらつき V_{Bz}

$$V_{Bz} = (\bar{z}_{[1]} - \bar{z}_{[-1]})^2$$

重み w の求め方

- ✓① 各クラスのサンプルは固まっている
 - z でのクラス内のばらつき V_{Wz}
- ✓② クラス1(赤)とクラス-1(青)は散らばっている
 - z でのクラス間のばらつき V_{Bz}

V_{Wz} が小さく(①)、 V_{Bz} が大きくなる(②) 直線を引く (w_1, w_2 を求める)

 $J = \frac{V_{Bz}}{V_{Wz}}$ が最大になる w_1, w_2 を求める

$$z = w_1 x_1 + w_2 x_2 = \mathbf{x} \mathbf{w}$$

$$\mathbf{x} = [x_1 \quad x_2], \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$J = \frac{(\bar{z}_{[1]} - \bar{z}_{[-1]})^2}{\sum_{i \in \text{クラス1}} (z^{(i)} - \bar{z}_{[1]})^2 + \sum_{i \in \text{クラス2}} (z^{(i)} - \bar{z}_{[-1]})^2}$$

$$= \frac{\mathbf{w}^T (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})^T (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]}) \mathbf{w}}{\mathbf{w}^T \left\{ \sum_{i \in \text{クラス1}} (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[1]})^T (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[1]}) + \sum_{i \in \text{クラス-1}} (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[-1]})^T (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[-1]}) \right\} \mathbf{w}}$$

$$= \frac{\mathbf{w}^T V_B \mathbf{w}}{\mathbf{w}^T V_W \mathbf{w}}$$

$\bar{\mathbf{x}}_{[k]}$: クラス k のみの
x の平均ベクトル

ただし、 $V_B = (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})^T (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})$

$$V_W = \sum_{i \in \text{クラス1}} (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[1]})^T (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[1]}) + \sum_{i \in \text{クラス-1}} (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[-1]})^T (\mathbf{x}^{(i)} - \bar{\mathbf{x}}_{[-1]})$$

w を求める

J が最大値 \Rightarrow J が極大値 \Rightarrow J を \mathbf{w} で偏微分して0

$$\frac{\partial J}{\partial \mathbf{w}} = \frac{2\{(\mathbf{w}^T V_B \mathbf{w}) V_W \mathbf{w} - (\mathbf{w}^T V_W \mathbf{w}) V_B \mathbf{w}\}}{(\mathbf{w}^T V_W \mathbf{w})^2} = 0$$

$$\underbrace{(\mathbf{w}^T V_B \mathbf{w})}_{\text{スカラー}} V_W \mathbf{w} - \underbrace{(\mathbf{w}^T V_W \mathbf{w})}_{\text{スカラー}} \underbrace{(\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})^T (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})}_{\text{スカラー}} \mathbf{w} = 0$$

$$V_W \mathbf{w} = \frac{(\mathbf{w}^T V_W \mathbf{w}) (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})^T \mathbf{w}}{(\mathbf{w}^T V_B \mathbf{w})} (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})$$

$\mathbf{w} \propto V_W^{-1} (\bar{\mathbf{x}}_{[1]} - \bar{\mathbf{x}}_{[-1]})$ \mathbf{w} の大きさは気にしなくてよい

クラス分類の結果の評価

✓混同行列 (confusion matrix)

		予測されたクラス	
		1 (Positive, 陽性)	-1 (Negative, 陰性)
実際の クラス	1 (Positive, 陽性)	True Positive (TP)	False Negative (FN)
	-1 (Negative, 陰性)	False Positive (FP)	True Negative (TN)

$$\text{正解率} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$\text{検出率} = \frac{TP}{TP + FN}$$

$$\text{精度} = \frac{TP}{TP + FP}$$

$$\text{誤検出率} = \frac{FP}{FP + TN} \quad \text{など}$$

クラス分類の結果の評価 例

✓混同行列 (confusion matrix)

		予測されたクラス	
		1 (Positive, 陽性)	-1 (Negative, 陰性)
実際の クラス	1 (Positive, 陽性)	45	5
	-1 (Negative, 陰性)	20	50

$$\text{正解率} = \frac{45 + 50}{45 + 5 + 20 + 50} = 0.79$$

$$\text{検出率} = \frac{45}{45 + 5} = 0.90$$

$$\text{精度} = \frac{45}{45 + 20} = 0.69$$

$$\text{誤検出率} = \frac{20}{20 + 50} = 0.29$$

(参考) Kappa係数

- ✓ 実際と予測結果の一致度を評価する指標
- ✓ Positive(陽性)データとNegative(陰性)データの偏りがある時に有効

$$\text{Kappa係数} = \frac{\text{正解率} - \text{偶然による一致率}}{1 - \text{偶然による一致率}}$$

$$\text{偶然による一致率} = \frac{\text{TP} + \text{FN}}{A} \times \frac{\text{TP} + \text{FP}}{A} + \frac{\text{FP} + \text{TN}}{A} \times \frac{\text{FN} + \text{TN}}{A}$$

$$(A = \text{TP} + \text{FN} + \text{FP} + \text{TN})$$

http://en.wikipedia.org/wiki/Cohen%27s_kappa

		予測されたクラス	
		1 (Positive, 陽性)	-1 (Negative, 陰性)
実際の クラス	1 (Positive, 陽性)	True Positive (TP)	False Negative (FN)
	-1 (Negative, 陰性)	False Positive (FP)	True Negative (TN)