

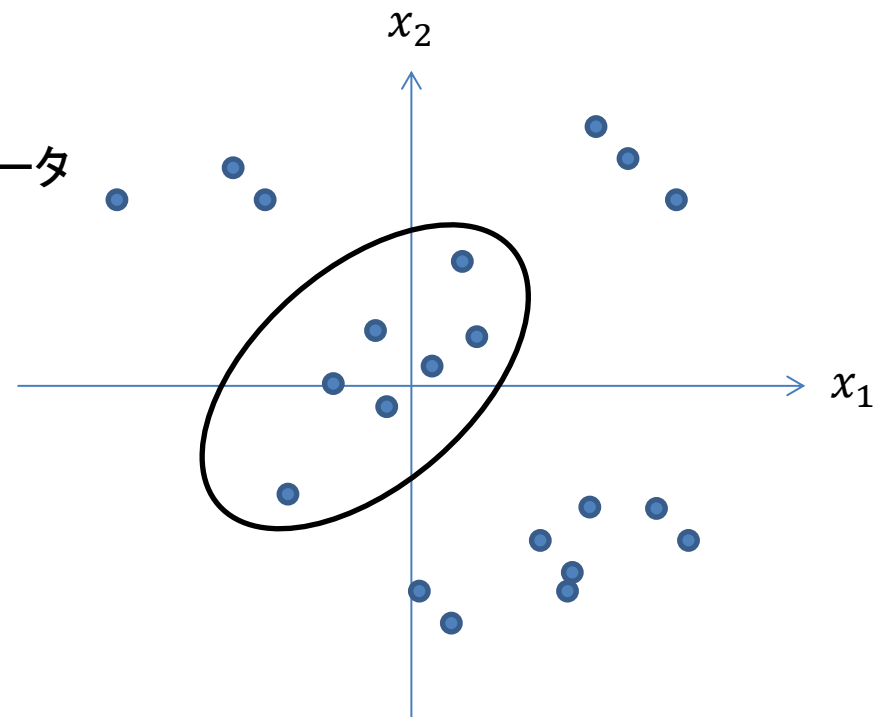
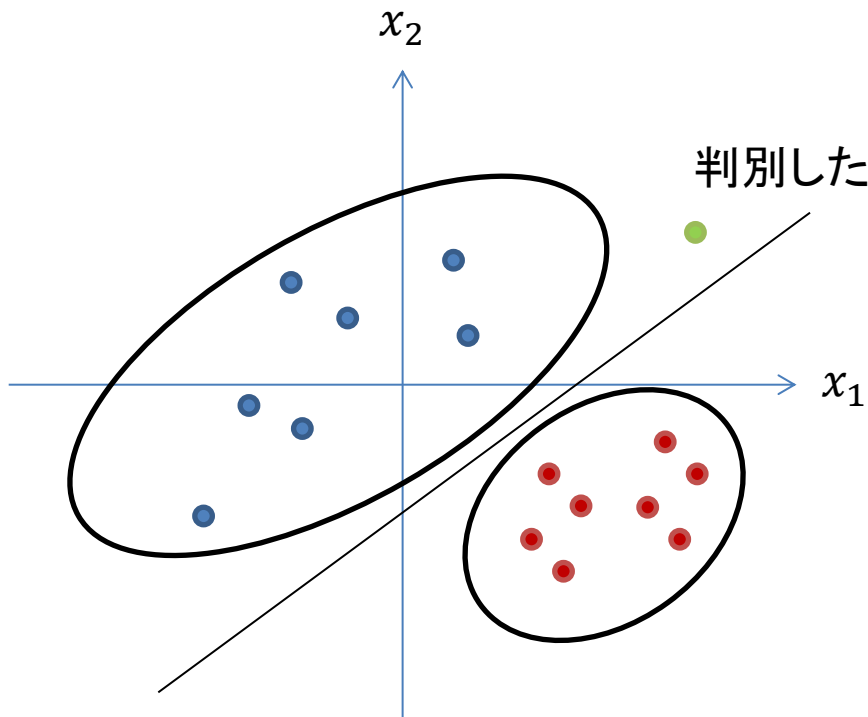
判別分析

- どちらのグループ(クラスター, 群)に属しているかが分かっているデータがあるときに, まだ分類されていないデータがどちらのグループに属するかを推定する方法.
- 一般に, 目的変数が質的変数, 説明変数が量的変数となる多変量解析.
- 説明変数が質的変数である場合もダミー変数を用いることで同様に扱うことができる.

判別分析の用語

- 目的変数
 - どちらのグループに属するかを示す変数.
 - 2グループであれば, $-1, 1$ 等と平均が0となるようにとる.
- 説明変数
 - 目的変数を説明変数の関数として定義する.
 - 説明変数は, 量的変数(連続値)であっても良いし, 質的変数(離散値)であっても良い. 質的変数の場合は, 整数値に変換する.

判別分析



重回帰分析による線形判別関数

1. 説明変数・目的変数を数値化する.
 - Yes/Noを判別するような場合, 目的変数は1,-1をとるようにする. (平均が0となるようにする.)
2. 回帰分析によって重回帰分析式を決定する.

$$y = a_0 + \sum a_i x_i$$

3. 目的変数 $y > 0$ の場合Yes, $y < 0$ の場合Noなどと判別する.

練習問題

健康/病気	因子1	因子2
健康	80	5
健康	60	3
病気	160	8
病気	140	6

因子1 = 100 ; 因子2 = 7のときはどちらか？

例題

- 質的変数を数値化する.
- 数値を平均値が0になるように変換する.

健康/病気	因子1	因子2
健康	80	5
健康	60	3
病気	160	8
病気	140	6

健康/病気	$(\text{因子1} - 100)/20$	$(\text{因子2} - 5)$
1	-1	0
1	-2	-2
-1	3	3
-1	2	1

演習問題1

- 新しい携帯電話についてアンケートをとったところ、以下のような結果となった。
- このとき、50歳男性と70歳女性が購入する可能性を判別せよ。

	年齢	性別	購入？
A	25	男	する
B	35	女	しない
C	70	男	しない
D	50	女	しない
E	30	女	しない
F	20	女	する
G	40	男	する

演習問題2

- 観測地点ごとの汚染/非汚染と汚染物資レベルが以下のようにになっている。
- このとき、 $SO_2=0.09$, $SPM=0.069$, $NOX=0.12$ の地点はどちらに判別されるか判定せよ。

汚染			非汚染		
SO2	SPM	NOX	SO2	SPM	NOX
0.01	0.078	0.21	0.017	0.078	0
0.015	0.073	0.11	0.013	0.047	0.05
0.011	0.072	0.09	0.008	0.082	0.04
0.01	0.042	0.06	0.03	0.021	0.01
0.025	0.113	0.23	0.05	0.04	0.02
0.015	0.064	0.09	0.06	0.035	0.03
0.12	0.06	0.11	0.008	0.045	0.14
0.016	0.075	0.08	0.009	0.074	0.04
0.024	0.092	0.2	0.007	0.039	0.05
0.025	0.074	0.14	0.007	0.059	0.05