§1 サポートベクターマシン（SVM）のアルゴリズム

サポートベクターマシン（SVM）は、1960年代に開発されたデータ識別用の技法で、現在でも広く利用されています。**マージンの最大化**というアイデアを用いて識別用の関数（すなわち**識別関数**）を求めます。一般論で調べると式が複雑になるので、具体例でそのしくみを見ることにしましょう。

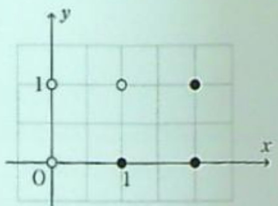
具体例で見てみよう

次の簡単な 例題 を通して、サポートベクターマシンの考え方を調べます。

例題 下表は、男性A、B、Cと女性D、E、Fを対象に、製品X、Yの好感度x、yを調べた結果です。この表から、SVMを用いて、男女を区別するx、yの線形の識別関数を求めましょう。

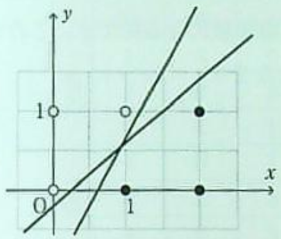
No	名前	好感度		類別
		x	y	
1	A	0	0	男
2	B	0	1	男
3	C	1	1	男
4	D	1	0	女
5	E	2	0	女
6	F	2	1	女

好感度(x, y)を座標とした点で、6人A～Fを表してみましょう。



白丸（○）は男、黒丸（●）は女。

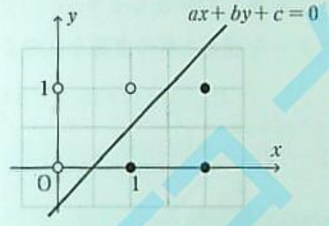
男女を識別する2変数x、yの「線形の識別関数」は、この平面上の直線を表します。ところで、白丸（男性）と黒丸（女性）を識別する直線はいく通りもあります（下図）。



白丸（○）と黒丸（●）を識別する直線はいろいろ引ける。

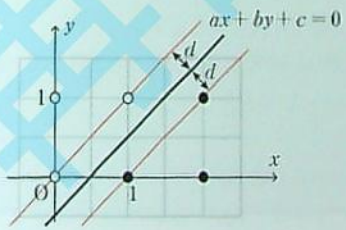
いく通りもある識別のための直線のうち、SVMは**マージンの最大化**という考え方で、その中の1本を確定します。この方程式を次のように表すことにし、今後「識別直線」と呼ぶことにします。

$$ax + by + c = 0 \quad (a、b\text{は同時に}0\text{にならない}) \cdots \boxed{1}$$



求めたい識別直線の方程式を
$ax+by+c=0$
とおく。

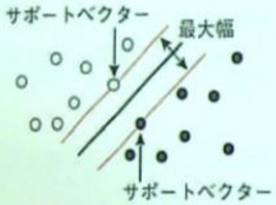
さて、この識別直線$\boxed{1}$と等間隔で、ちょうど男女の境界の縁（マージン）を通る平行な2直線を描いてみましょう。



識別直線$\boxed{1}$に平行に2直線を引く。各々は識別するデータの縁（マージン）を通るとする。このとき、幅dを最大化することが「マージンの最大化」という考え方である。

「マージンの最大化」とは、この平行な2直線の間隔dを最大にすることです。そして、この縁にある男女のデータ要素(最低各1つ以上)を**サポートベクター**といいます。

直感的に言うと、男女のデータを区切る最大幅の片側1車線直線道路を作ったとき、そのセンターラインが求めたい識別関数になるのです。このとき、路肩にある男女の点が「サポートベクター」となります。



最大幅の片側1車線直線道路でデータを分けるイメージがSVMの原理。そのセンターラインが目的の識別関数を表す。

マージンの最大化を式で表現

では、この「マージンの最大化」を式で表現してみましょう。

識別直線$\boxed{1}$に平行で、サポートベクターを通る2本の直線を次のように置いてみましょう。

$$ax+by+c=1,\ ax+by+c=-1 \cdots \boxed{2}$$

係数a、b、cは符号を含めた定数倍の不定性を持っているので、このように置いても一般性を欠くことはありません(▶下記MEMO参照)。

MEMO 直線の式の不定性

1本の直線を表す方程式$ax+by+c=0$には不定性があります。
例えば、方程式$2x+3y+4=0$が表す直線は、次のように表された直線と同一です。

$$4x+6y+8=0,\ -2x-3y-4=0$$

そこで、式$\boxed{2}$のように直線を置くことが許されるわけです。

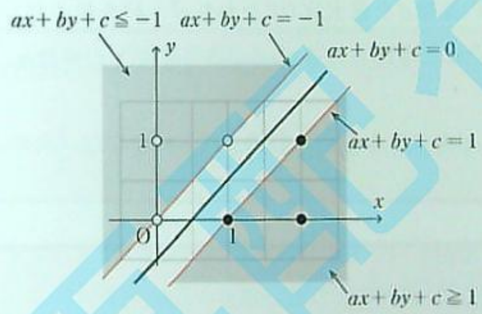
さて、与えられた資料において、男に所属するデータの要素には-1を与え「負例」と呼ぶことにします。また、女に所属するデータの要素には1を与え、「正例」と呼ぶことにします。

	No	名	x	y	類別	正負
負例	1	A	0	0	男	-1
	2	B	0	1	男	-1
	3	C	1	1	男	-1
正例	4	D	1	0	女	1
	5	E	2	0	女	1
	6	F	2	1	女	1

注 女を負例にし、男を正例にしても、結論は同じ。

ここで、次の関係を仮定しましょう。i番目の人の好感度(x, y)を(x_i, y_i)として、

$$\left.\begin{array}{l}\text{負例(男)}:ax_i+by_i+c\leqq -1\\ \text{正例(女)}:ax_i+by_i+c\geqq 1\end{array}\right\}\cdots\boxed{3}$$



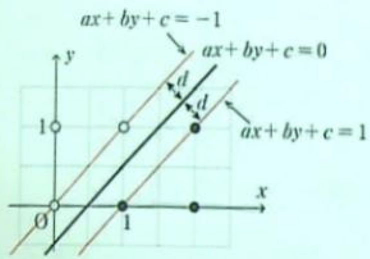
3次元で考えると、平面
$z=ax+by+c$
の下側に○、上側に●が配置されることになる。

ここからはサポートベクトルに焦点を当てます。サポートベクトルとなるデータ要素を表す点(x_i, y_i)は、式$\boxed{2}$から次の関係を満たします。

$$\left.\begin{array}{l}\text{負例(すなわち男)}:ax_i+by_i+c=-1\\ \text{正例(すなわち女)}:ax_i+by_i+c=1\end{array}\right\}\cdots\boxed{4}$$

サポートベクターから識別直線$\boxed{1}$に引いた垂線の長さをdとしましょう。すると、高校で習う「点と直線の距離の公式」から、サポートベクターとなるデータの点(x_i, y_i)から識別直線$\boxed{1}$に引いた垂線の長さは次のように表せます。

$$d=\frac{|ax_i+by_i+c|}{\sqrt{a^2+b^2}}$$



マージン最大化の方針のために、幅dは最大化する。

サポートベクターの座標(x_i, y_i)は式$\boxed{4}$を満たすので次式が成立します。

$$d=\frac{1}{\sqrt{a^2+b^2}} \cdots \boxed{5}$$

マージン最大化の方針は、この距離dを最大化(すなわちa^2+b^2を最小化)することを要求します。そして、そのとき得られるa、bが識別直線$\boxed{1}$の係数a、bになるわけです。

こうして、直線$\boxed{1}$とサポートベクターを求める目標が得られました。次のようにまとめられます。

> データを表す点(x_i, y_i)について$(i=1, 2, \cdots, 6)$、
>
> $$\left.\begin{array}{l}\text{負例に対して}\ ax_i+by_i+c \leqq -1,\\ \text{正例に対して}\ ax_i+by_i+c \geqq 1\end{array}\right\} \cdots \boxed{3}\ (\text{再掲})$$
>
> の条件の下で、次の式の値を最小にするa、b、cを求める。
>
> $$\frac{1}{2}(a^2+b^2) \cdots \boxed{6}$$

こうして得られたa、b、cについて、条件$\boxed{4}$を満たす点(x_i, y_i)がサポートベクトルになります。ちなみに、式$\boxed{6}$の係数$\frac{1}{2}$は今後の計算が見やすいようにするために付加しました。「a^2+b^2の最小化」とそれを半分にした式$\boxed{6}$の最小化とは同じだからです。

双対問題に変換

ここで突然ですが、話を一般化できるように、次のような変数t_iを導入します$(i=1, 2, \cdots, 6)$。

正例に対して$t_i=1$、負例にたいして、$t_i=-1$

すると、簡単な計算から、条件$\boxed{3}$は次のように1つにまとめられます。

$$t_i(ax_i+by_i+c) \geqq 1 \cdots \boxed{7}$$

注 tはteacherの頭文字。SVMにおいて、この値t_iが正解ラベルになります。

こうして、識別直線$\boxed{1}$の係数a、bを求める条件$\boxed{3}$が1つに表現されました。すると目標はさらに次のように簡潔に表現できます。

> $t_i(ax_i+by_i+c) \geqq 1$の条件で、
>
> $\frac{1}{2}(a^2+b^2)$を最小にするa、b、cを求める。$\cdots \boxed{8}$

さて、▶2章§3では、「双対問題」という技法を調べました。これを利用すると、この目標$\boxed{8}$は次の双対問題に置き換えられます。

> $$L=\frac{1}{2}(a^2+b^2)+\mu_1\{1-t_1(ax_1+by_1+c)\}+\mu_2\{1-t_2(ax_2+by_2+c)\}$$
> $$+\cdots+\mu_6\{1-t_6(ax_6+by_6+c)\} \cdots \boxed{9}$$
>
> について、a、b、cに関する最小値を求める。さらに、得られたμ_1、$\cdots$、μ_6の式について、その最大値を求める。ただし、μ_1、$\cdots$、μ_6は0以上。

では、この双対問題を具体的に解いてみましょう。

最初に、a、b、cに関して、このLの最小値mを求めてみます。それには次のように偏微分が利用できます(▶付録E)。

$$\frac{\partial L}{\partial a}=a-\mu_1t_1x_1-\mu_2t_2x_2-\cdots-\mu_6t_6x_6=0$$
$$\frac{\partial L}{\partial b}=b-\mu_1t_1y_1-\mu_2t_2y_2-\cdots-\mu_6t_6y_6=0$$
$$\frac{\partial L}{\partial c}=-t_1\mu_1-t_2\mu_2-\cdots-t_6\mu_6=0$$

これから、次の関係式が得られます。

$$\left.\begin{aligned}a&=\mu_1t_1x_1+\mu_2t_2x_2+\cdots+\mu_6t_6x_6\\ b&=\mu_1t_1y_1+\mu_2t_2y_2+\cdots+\mu_6t_6y_6\end{aligned}\right\}\cdots\boxed{10}$$

$$t_1\mu_1+t_2\mu_2+\cdots+t_6\mu_6=0\cdots\boxed{11}$$

式$\boxed{9}$のLに代入して、

$$\begin{aligned}L&=\frac{1}{2}(a^2+b^2)+\mu_1\{1-t_1(ax_1+by_1+c)\}+\mu_2\{1-t_2(ax_2+by_2+c)\}\\&\qquad+\cdots+\mu_6\{1-t_6(ax_6+by_6+c)\}\\&=\frac{1}{2}(a^2+b^2)-a(\mu_1t_1x_1+\mu_2t_2x_2+\cdots+\mu_6t_6x_6)\\&\quad-b(\mu_1t_1y_1+\mu_2t_2y_2+\cdots+\mu_6t_6y_6)+(\mu_1+\mu_2+\cdots+\mu_6)\\&=\frac{1}{2}(a^2+b^2)-a(a)-b(b)+(\mu_1+\mu_2+\cdots+\mu_6)\end{aligned}$$

これを整理して、最小化すべき目標の式$\boxed{9}$は次のように簡潔にまとめられます。

$$L=-\frac{1}{2}(a^2+b^2)+(\mu_1+\mu_2+\cdots+\mu_6)\cdots\boxed{12}$$

ここで、a、bは式$\boxed{10}$で与えられます。

▶計算しやすいように変形

さらに式を変形しましょう。式$\boxed{10}$より、

$$\begin{aligned}a^2&=(\mu_1t_1x_1+\mu_2t_2x_2+\cdots+\mu_6t_6x_6)^2\\&=\mu_1\mu_1t_1t_1x_1x_1+\mu_1\mu_2t_1t_2x_1x_2+\mu_1\mu_3t_1t_3x_1x_3+\cdots+\mu_6\mu_6t_6t_6x_6x_6\\ b^2&=(\mu_1t_1y_1+\mu_2t_2y_2+\cdots+\mu_6t_6y_6)^2\\&=\mu_1\mu_1t_1t_1y_1y_1+\mu_1\mu_2t_1t_2y_1y_2+\mu_1\mu_3t_1t_3y_1y_3+\cdots+\mu_6\mu_6t_6t_6y_6y_6\end{aligned}$$

こうして、式$\boxed{12}$のLは与えられたデータx_i、y_i、t_i $(i=1,2,\cdots,6)$と求めたい変数μ_1、μ_2、…、μ_6だけの式に変形されました。

$$\begin{aligned}L=-\frac{1}{2}\{&\mu_1\mu_1t_1t_1(x_1x_1+y_1y_1)+\mu_1\mu_2t_1t_2(x_1x_2+y_1y_2)\\&+\mu_1\mu_3t_1t_3(x_1x_3+y_1y_3)+\cdots+\mu_6\mu_6t_6t_6(x_6x_6+y_6y_6)\}\cdots\boxed{13}\\&+(\mu_1+\mu_2+\cdots+\mu_6)\end{aligned}$$

このように表現すれば、計算がしやすく、一般的なデータへの拡張も容易でしょう。すなわち、目標は次のように定まったのです。

> 条件$\boxed{11}$の下で、0以上のμ_1、μ_2、…、μ_6について、式$\boxed{13}$のLの最大値を求める。…$\boxed{14}$

注 データの大きさの数だけある制約条件$\boxed{7}$が、たった1個の制約条件$\boxed{11}$に激減したのです。プログラム作成上、これはありがたい結果です。

この目標$\boxed{14}$はExcelの得意とするところです。次の節で、実際に計算してみましょう。その結果は次の通りです。

$$\left.\begin{aligned}&\mu_1=1.648、\mu_2=0.000、\mu_3=2.352\\&\mu_4=3.648、\mu_5=0.000、\mu_6=0.352\end{aligned}\right\}\cdots\boxed{15}$$

これから、式$\boxed{10}$を利用して、

$$a=2、b=-2\cdots\boxed{16}$$

サポートベクターと定数項cを求める

サポートベクターを求めましょう。式4から、与えられたデータ要素の中で、サポートベクターとなるデータ(x, y)は次の式を満たします。

$$\left.\begin{array}{ll}(\text{正例})\ ax+by+c=1 & \text{すなわち、}c=1-(ax+by)\\ (\text{負例})\ ax+by+c=-1 & \text{すなわち、}c=-1-(ax+by)\end{array}\right\}\cdots\boxed{17}$$

ところで、ラグランジュ双対問題の式9からわかるように、「a^2+b^2の最小化」に関与するものは$\mu_i>0$の場合です(関与しなければ$\mu_i=0$)。このことと、上記17に結果15 16を代入して、次の表が作成できます。

	No	名	x	y	μ	SV	c
負例	1	A	0	0	1.648	YES	−1.000
	2	B	0	1	0.000	NO	
	3	C	1	1	2.352	YES	−1.000
正例	4	D	1	0	3.648	YES	−1.000
	5	E	2	0	0.000	NO	
	6	F	2	1	0.352	YES	−1.000

注 表頭にある「SV」はサポートベクターの略。

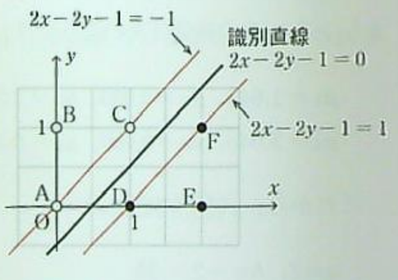
この表から識別直線1の定数項cが定められました。

$c=-1\cdots\boxed{18}$

目的の識別直線の方程式1は式16 18から次のように得られます。

$2x-2y-1=0\cdots\boxed{19}$

これまでに得られた解が条件を満たしていることを、右図で確かめてください。



識別直線の方程式は19となる。データの中で、要素A、C、D、Fがサポートベクターであることがわかる。

§2 サポートベクターマシン(SVM)をExcelで体験

前節で調べたSVMのしくみをExcelで確認してみましょう。

ExcelでSVM

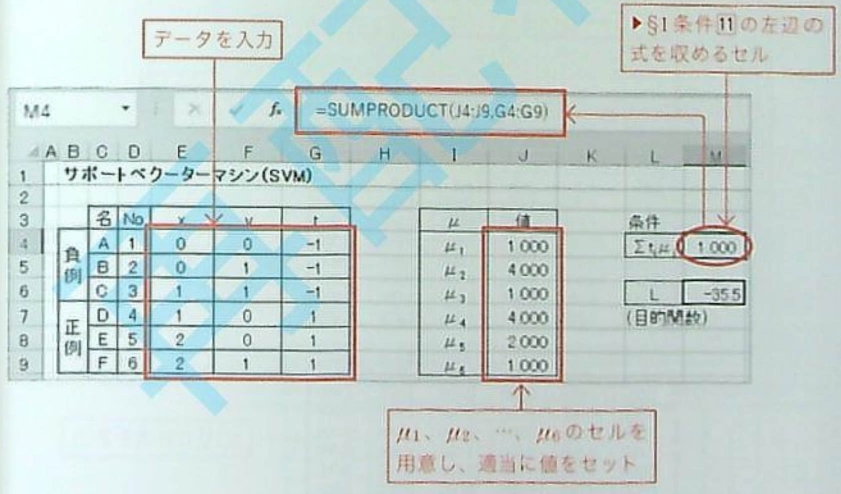
前節で調べた例題を、ステップを追って調べましょう。

演習 ▶§1の例題の解を、Excelで求めましょう。

注 本節のワークシートは、ダウンロードサイト(▶244ページ)に掲載されたファイル「4.xlsx」にあります。

① データを入力し、μ_1、μ_2、…、μ_6のセルを用意し値をセットします。

μ_1、μ_2、…、μ_6には適当な値を設定します。また、それらの満たすべき条件式(▶§1の11)を収めるセルも用意します。

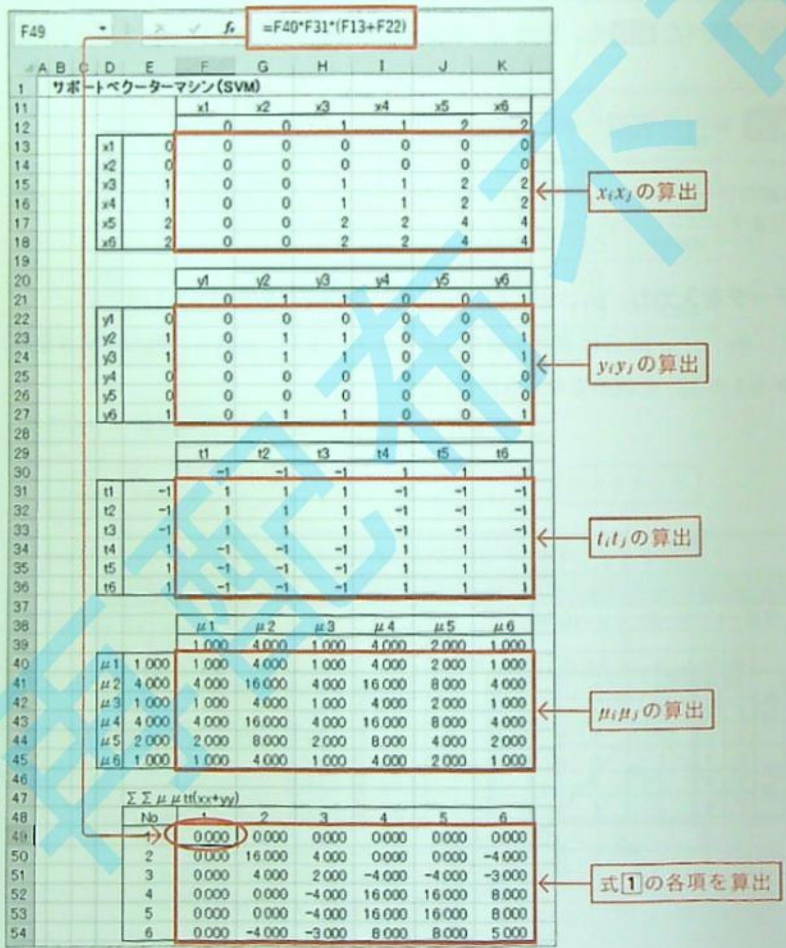


② ▶§1の式13のLを算出する準備をします。

▶§1の式13 $(=L)$ の中括弧{ }の中の式(下記1)を算出するために、表を作成します。

$$\{\mu_1\mu_1 t_1 t_1(x_1x_1+y_1y_1)+\mu_1\mu_2 t_1 t_2(x_1x_2+y_1y_2)+\mu_1\mu_3 t_1 t_3(x_1x_3+y_1y_3)+\cdots+\mu_6\mu_6 t_6 t_6(x_6x_6+y_6y_6)\} \quad \cdots \boxed{1}$$

下記のワークシートでは、この式1の各項を表(タイトル名「ΣΣμμtt(xx+y)」)に求めています。



		x1	x2	x3	x4	x5	x6
		0	0	1	1	2	2
x1	0	0	0	0	0	0	0
x2	0	0	0	0	0	0	0
x3	1	0	0	1	1	2	2
x4	1	0	0	1	1	2	2
x5	2	0	0	2	2	4	4
x6	2	0	0	2	2	4	4

		y1	y2	y3	y4	y5	y6
		0	1	1	0	0	1
y1	0	0	0	0	0	0	0
y2	1	0	1	1	0	0	1
y3	1	0	1	1	0	0	1
y4	0	0	0	0	0	0	0
y5	0	0	0	0	0	0	0
y6	1	0	1	1	0	0	1

		t1	t2	t3	t4	t5	t6
		−1	−1	−1	1	1	1
t1	−1	1	1	1	−1	−1	−1
t2	−1	1	1	1	−1	−1	−1
t3	−1	1	1	1	−1	−1	−1
t4	1	−1	−1	−1	1	1	1
t5	1	−1	−1	−1	1	1	1
t6	1	−1	−1	−1	1	1	1

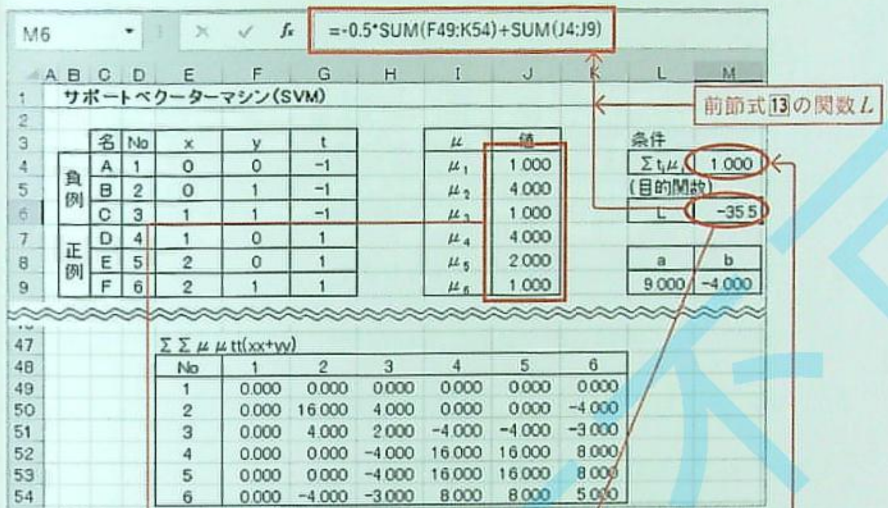
		μ1	μ2	μ3	μ4	μ5	μ6
		1.000	4.000	1.000	4.000	2.000	1.000
μ1	1.000	1.000	4.000	1.000	4.000	2.000	1.000
μ2	4.000	4.000	16.000	4.000	16.000	8.000	4.000
μ3	1.000	1.000	4.000	1.000	4.000	2.000	1.000
μ4	4.000	4.000	16.000	4.000	16.000	8.000	4.000
μ5	2.000	2.000	8.000	2.000	8.000	4.000	2.000
μ6	1.000	1.000	4.000	1.000	4.000	2.000	1.000

ΣΣμμtt(xx+yy)

No	1	2	3	4	5	6
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	16.000	4.000	0.000	0.000	−4.000
3	0.000	4.000	2.000	−4.000	−4.000	−3.000
4	0.000	0.000	−4.000	16.000	16.000	8.000
5	0.000	0.000	−4.000	16.000	16.000	8.000
6	0.000	−4.000	−3.000	8.000	8.000	5.000

③ Lを求め、Excelアドインのソルバーにセットします。

▶§1の式13 $(=L)$ を計算するセルを用意し、次の図のようにソルバーに設定します。



	名	No	x	y	t
負例	A	1	0	0	−1
	B	2	0	1	−1
	C	3	1	1	−1
正例	D	4	1	0	1
	E	5	2	0	1
	F	6	2	1	1

μ	値
μ_1	1.000
μ_2	4.000
μ_3	1.000
μ_4	4.000
μ_5	2.000
μ_6	1.000

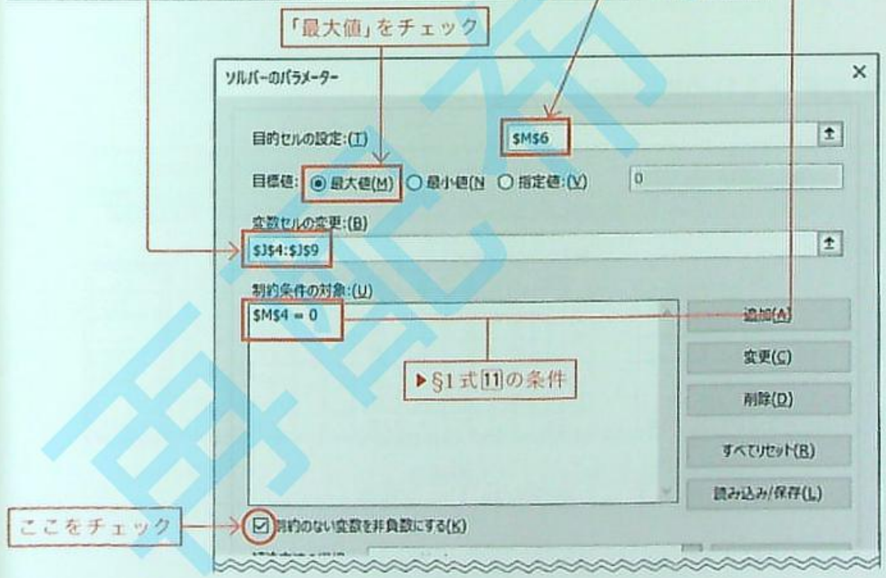
条件

$\Sigma t_i\mu_i$	1.000
(目的関数)	
L	−35.5

a	b
9.000	−4.000

ΣΣμμtt(xx+yy)

No	1	2	3	4	5	6
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	16.000	4.000	0.000	0.000	−4.000
3	0.000	4.000	2.000	−4.000	−4.000	−3.000
4	0.000	0.000	−4.000	16.000	16.000	8.000
5	0.000	0.000	−4.000	16.000	16.000	8.000
6	0.000	−4.000	−3.000	8.000	8.000	5.000



④ ソルバーの出力結果を示しましょう。

L9　=SUMPRODUCT(J4:J9,G4:G9,E4:E9)

サポートベクターマシン（SVM）

	名	No	x	y	t
負例	A	1	0	0	−1
	B	2	0	1	−1
	C	3	1	1	−1
正例	D	4	1	0	1
	E	5	2	0	1
	F	6	2	1	1

最適化の結果

μ	値
μ_1	1.648
μ_2	0.000
μ_3	2.352
μ_4	3.648
μ_5	0.000
μ_6	0.352

条件	
$\Sigma t_i\mu_i$	0.000
（目的関数）	
L	4

▶§1 式16

a	b
2.000	−2.000

こうして、前節（▶§1）の式15の値が得られます。

$$\mu_1 = 1.648、\mu_2 = 0.000、\mu_3 = 2.352、\mu_4 = 3.648、\mu_5 = 0.000、\mu_6 = 0.352$$

また、前節（▶§1）の式10から、方程式の係数a、bが求められます。

$$\left.\begin{array}{l} a = \mu_1 t_1 x_1 + \mu_2 t_2 x_2 + \cdots + \mu_6 t_6 x_6 = 2 \\ b = \mu_1 t_1 y_1 + \mu_2 t_2 y_2 + \cdots + \mu_6 t_6 y_6 = -2 \end{array}\right\} \cdots$$ ▶§1の16

⑤ 前節（▶§1）の式17からcを算出します。

X4　=IF(W4="YES",U4-SUMPRODUCT($L$9:$M$9,R4:S4),"")

条件	
$\Sigma t_i\mu_i$	0.000
（目的関数）	
L	4

a	b
2.000	−2.000

●まとめとcの算出

cの算出

	No	名	x	y	類別	正負(t)	μ	SV	c
負例	1	A	0	0	男	−1	1.648	YES	−1.000
	2	B	0	1	男	−1	0.000	NO	
	3	C	1	1	男	−1	2.352	YES	−1.000
正例	4	D	1	0	女	1	3.648	YES	−1.000
	5	E	2	0	女	1	0.000	NO	
	6	F	2	1	女	1	0.352	YES	−1.000

こうして、前節（▶§1）の識別直線1の定数項cが定められました。

$$c = -1$$

以上から、識別直線の方程式（前節（▶§1）の19）が得られます。

5章 ニューラルネットワークとディープラーニング

ニューラルネットワークは近年のAIブームの火付け役になったAIの基本モデルです。そこで利用されている誤差逆伝播法は、AIの多くの分野で利用されています。しくみを調べてみましょう。

（注）本章▶§1以外では、ニューラルネットワークという言葉を、ディープラーニングを含む広い意味で利用しています。