

多次元尺度法と単調回帰

前川眞一

東京工業大学 大学院社会理工学研究科 970410, 071023,15,26,27,30 *

個体間の類似度が与えられた時に、似ているものは近くに、似ていないものは遠くに、という具合に類似度を距離に変換し、個体間の距離を最も良く再現するような個体の座標を求める方法を多次元尺度法 (Multi-Dimensional Scaling, MDS) と呼ぶ。MDS には、解を解析的に求めようとする古典的多次元尺度法 (Classical MDS) と、データの最適変換を含み、データに距離を直接最小 2 乗法を用いて当てはめる非計量的多次元尺度法 (non-metric MDS) とに大別される。

1 古典的多次元尺度法

n 個の個体の r 次元の座標行列 X から計算される個体間の 2 乗距離

$$d_{ij}^2 = \sum_{a=1}^r (x_{ia} - x_{ja})^2, i, j = 1, 2, \dots, n \quad (1)$$

を要素として持つ $n \times n$ の行列を D^2 とする。これは、行列を用いて書くと

$$D^2 = \text{diag}(XX')\mathbf{1}\mathbf{1}' - 2XX' + \mathbf{1}\mathbf{1}'\text{diag}(XX') \quad (2)$$

と書ける。ただし、 $\mathbf{1}$ は要素が全て 1 である $n \times 1$ のベクトルである。

いま、中心化行列を

$$J = I - \frac{1}{n}\mathbf{1}\mathbf{1}' \quad (\text{したがって } J = J', J\mathbf{1} = \mathbf{0}) \quad (3)$$

と定義すれば、

$$J = J' \quad \text{and} \quad J\mathbf{1} = \mathbf{0}, \quad \text{and} \quad \mathbf{1}'J = \mathbf{0}' \quad (4)$$

であり、これを用いれば、中心化された座標行列 Y は

$$Y = JX \quad \text{or} \quad y_{ia} = x_{ia} - \frac{1}{n} \sum_{k=1}^n x_{ka} \quad (5)$$

と書ける。したがって、この行列を D^2 行列の左右から掛け、それを $-\frac{1}{2}$ 倍すれば、

$$B = -\frac{1}{2}JD^2J \quad (6)$$

$$= -\frac{1}{2}J(\text{diag}(XX')\mathbf{1}\mathbf{1}' - 2XX' + \mathbf{1}\mathbf{1}'\text{diag}(XX'))J \quad (7)$$

$$= JXX'J = YY' \quad (8)$$

* 〒 152-8552 目黒区大岡山 2-12-1

Telefax: 81-3-5734-3242, e-mail: mayekawa@hum.titech.ac.jp

本稿第 4 節は前川 (1997) の第 4.3.2.2 節に加筆修正したものである。

となるが、この操作を D^2 行列の2重中心化と呼び、得られる B 行列をスカラー積行列と呼ぶ。なお、 B 行列の要素は、以下のようにになっている。

$$\begin{aligned}
b_{ij} &= -\frac{1}{2} \left(d_{ij}^2 - \frac{1}{n} \sum_{j=1}^n d_{ij}^2 - \frac{1}{n} \sum_{i=1}^n d_{ij}^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 \right) \\
&= \sum_{a=1}^r x_{ia} x_{ja} - \frac{1}{n} \sum_{a=1}^r x_{ia} \sum_{j=1}^n x_{ja} - \frac{1}{n} \sum_{a=1}^r x_{ja} \sum_{i=1}^n x_{ia} + \frac{1}{n^2} \sum_{a=1}^r \left(\sum_{i=1}^n x_{ia} \right) \left(\sum_{j=1}^n x_{ja} \right) \\
&= \sum_{a=1}^r \left(x_{ia} - \frac{1}{n} \sum_{k=1}^n x_{ka} \right) \left(x_{ja} - \frac{1}{n} \sum_{k=1}^n x_{ka} \right) \\
&= \sum_{a=1}^r y_{ia} y_{ja}
\end{aligned}$$

これより、 B 行列の固有値分解を

$$B = P \Lambda P' \quad (9)$$

とすれば、

$$Y = P \left(\Lambda^{\frac{1}{2}} \right) \quad (10)$$

となる。ただし、原点 c と正規直交行列 T に関して不定性が残る。すなわち、

$$X = Y T + \mathbf{1} c' \quad (11)$$

したがって、 o_{ij} , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$ を個体 i, j 間の観測された距離（非類似度）とし、それから計算されるスカラー積行列を B とすれば、

$$RSS = \sum_{i=1}^n \sum_{j=1}^n (b_{ij} - \sum_{a=1}^r y_{ia} y_{ja})^2 \quad (12)$$

を最小とする Y を求めればよい。その r 次元の解は、 B 行列の固有値分解を

$$B \approx P_r \Lambda_r P_r' \quad (13)$$

とすれば、

$$Y = P_r \left(\Lambda_r^{\frac{1}{2}} \right) \quad (14)$$

となる。

2 最小2乗解と非計量的多次元尺度法

o_{ij} , $i = 2, 3, \dots, n$, $j = 1, 2, \dots, i$ を個体 i, j 間の観測された距離（非類似度）とする。観測値にモデルをフィットする場合、以下のような最小2乗基準を最小とするような X 行列を直接求めることが出来る。（数値計算、最適化）

$$RSS1 = \sum_{i=2}^n \sum_{j=1}^i (o_{ij} - d_{ij})^2 \quad \text{or} \quad RSS2 = \sum_{i=2}^n \sum_{j=1}^i (o_{ij} - d_{ij}^2)^2 \quad (15)$$

また、観測値を変換したものに距離をフィットする場合は、データの最適変換（最小2乗基準を最小とするような変換）を $\tau(\cdot)$ として、

$$\text{RSS1} = \sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - d_{ij})^2 \quad \text{or} \quad \text{RSS2} = \sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - d_{ij}^2)^2 \quad (16)$$

を最小とするような \mathbf{X} 行列と最適変換を同時に求めることが出来る。なお、最適変換としては、

- パラメトリックな変換
 $\tau(o) = ao + b$, $\tau(o) = a \log o + b$, $\tau(o) = ao^b + c$ 等
- 単調増加変換
 $o_{ij} \leq o_{kl}$ ならば $\tau_{ij}(\mathbf{o}) \leq \tau(o_{kl})$

等が考えられる。単調変換（データが類似度の場合は単調減少関数、非類似度の場合は単調増加関数）を用いる場合は、非計量的多次元尺度法 (non-metric MDS) と呼ばれている。なお、最適変換としてはこれ以外にも単調増加スプラインを用いる方法（前川,1997）や単調増加ニューラルネットワークを用いる方法（van Wezel & Kosters, 2004）等がある。

パラメタの推定に際しては、交互最小2乗法 (Alternating Least Squares, ALS) を用いる事が多く、最適変換を所与として RSS を最小とする座標 \mathbf{X} を求める、座標を所与として最適変換を求める、というふたつのステップを繰り返していく。ガウス・ニュートン法を用いるとして必要となる偏微分は

$$\begin{aligned} \frac{\partial d_{kj}}{\partial x_{ka}} &= \frac{(x_{ka} - x_{ja})}{d_{kj}} \\ \frac{\partial d_{jk}}{\partial x_{ka}} &= -\frac{(x_{ka} - x_{ja})}{d_{kj}} \end{aligned}$$

また、それ以外の偏微分（たとえば、 \mathbf{x}_k の要素を含まない d_{ij} , $i, j \neq k$ の \mathbf{x}_k に関する偏微分は全てゼロである。したがって

$$\frac{\partial \text{RSS}}{\partial x_{ka}} = -2 \sum_{j=1}^n (o_{kj} - d_{kj}) \frac{(x_{ka} - x_{ja})}{d_{kj}} = 0$$

を満たす \mathbf{X} を探すことになる。また、最適変換に含まれるパラメタを ζ と書くと、それに関しても、

$$\frac{\partial \text{RSS}}{\partial \zeta} = 0 \quad (17)$$

となっていなければならない。

なお、最適変換を含む場合、何らかの制約を入れないと RSS1 や RSS2 はどんどん小さくなってしまふ。そこで、それを防ぐために、変換された距離の分散で基準化した残差

$$\text{RSS1} = \frac{\sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - d_{ij})^2}{\sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - \overline{\tau_{ij}(\mathbf{o})})^2} \quad \text{or} \quad \text{RSS2} = \frac{\sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - d_{ij}^2)^2}{\sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - \overline{\tau_{ij}(\mathbf{o})})^2} \quad (18)$$

を最小としなければならないが、これは、上記の繰り返し計算の中で、最適変換の値を、その分散が 1 となるように基準化する事によって達成される。詳細は前川 (1997) の第 4.5 節や 5.2 節を参照されたい。

3 個人差モデル

$o_{ijk}, i = 2, 3, \dots, n, j = 1, 2, \dots, i, k = 1, 2, \dots, N$ を、個人 k による個体 i, j 間の距離（非類似度）とする。個人 k は共通座標 \mathbf{X} の各軸を w_{ka} で重み付けして（非）類似度を評価していると仮定する。図 1 を参照。

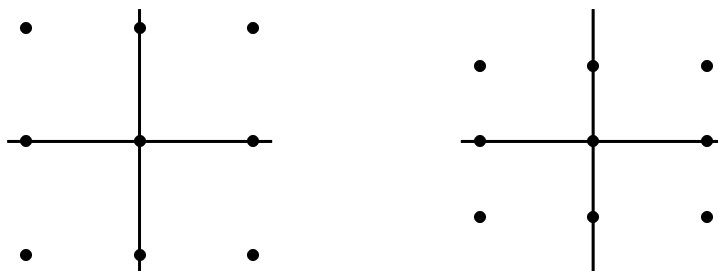


図 1: 個人差 MDS モデル
 $w_{s1} > w_{s2}$ の場合

重み付きユークリッド距離

$$d_{ijk}^2 = \sum_{a=1}^r w_{ka} (x_{ia} - x_{ja})^2 \quad (19)$$

モデルフィットの度合いとしては個人ごとの RSS を加えたもの

$$\text{RSS1} = \sum_{k=1}^S \sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - d_{ijk})^2 \quad \text{or} \quad \text{RSS2} = \sum_{k=1}^S \sum_{i=2}^n \sum_{j=1}^i (\tau_{ij}(\mathbf{o}) - d_{ijk}^2)^2 \quad (20)$$

を用いる。

パラメタ推定に際してガウス・ニュートン法を用いるとして必要となる偏微分は

$$\begin{aligned} \frac{\partial d_{kjs}}{\partial x_{ka}} &= \frac{w_{sa}(x_{ka} - x_{ja})}{d_{kjs}} \\ \frac{\partial d_{jks}}{\partial x_{ka}} &= -\frac{w_{sa}(x_{ka} - x_{ja})}{d_{kjs}} \\ \frac{\partial d_{ijm}}{\partial w_{sa}} &= \frac{(x_{ia} - x_{ja})^2}{2d_{ijs}} \end{aligned}$$

また、それ以外の偏微分（たとえば、 \mathbf{x}_k の要素を含まない $d_{ijs}, i, j \neq k$ の \mathbf{x}_k に関する偏微分、 \mathbf{w}_s の要素を含まない $d_{ijm}, m \neq s$ の \mathbf{w}_s に関する偏微分は全てゼロである。したがって、

$$\begin{aligned} \frac{\partial \text{RSS}}{\partial x_{ka}} &= -2 \sum_{s=1}^S \sum_{j=1}^n (o_{kjs} - d_{kjs}) \frac{w_{sa}(x_{ka} - x_{ja})}{d_{kjs}} \\ \frac{\partial \text{RSS}}{\partial w_{sa}} &= -\sum_{i>j}^n (o_{ijs} - d_{ijs}) \frac{(x_{ia} - x_{ja})^2}{d_{ijs}} \end{aligned}$$

となる。

4 最小2乗単調回帰

ALS を用いて非計量的 MDS の計算をする場合、 \mathbf{X} したがって \mathbf{D} が与えられたときに

$$\text{RSS} = \sum_{i>j}^n (d_{ij} - \tau_{ij}(\mathbf{o}|\boldsymbol{\xi}))^2 \quad (21)$$

を最小とし、条件

$$o_{ij} \leq o_{kl} \rightarrow \tau_{ij}(\mathbf{o}|\boldsymbol{\xi}) \leq \tau_{kl}(\mathbf{o}|\boldsymbol{\xi}) \quad (22)$$

を満たす単調変換 $\tau(\boldsymbol{\xi})$ を求める必要があるが、これは基準変数 d_{ij} を順序尺度の説明変数 o_{ij} に回帰していると考えることが出来る。

より一般的には、間隔尺度以上の水準の基準変数を v 、順序尺度の説明変数を u と記して、 (v_i, u_i) , $i = 1, 2, \dots, N$ が観測されている場合に、 v を u に回帰することを考える。

すなわち、以下の単調関係

$$u_i \leq u_j \rightarrow \tau_i(\mathbf{u}) \leq \tau_j(\mathbf{u}) \quad (23)$$

を満たす \mathbf{u} の関数をまとめて $\tau(\mathbf{u}|\boldsymbol{\xi})$ と記す。ただし $\boldsymbol{\xi}$ は変換のパラメタである。このような変換は無数にあるが、基準変数 v が与えられた場合に、モデル

$$v_i = \tau_i(\mathbf{u}|\boldsymbol{\xi}) + \varepsilon_i \quad (24)$$

を考え、

$$\text{RSS} = \sum_{i=1}^N (v_i - \tau_i(\mathbf{u}|\boldsymbol{\xi}))^2$$

を最小とするものは、 v の u への最小2乗単調回帰 (monotone least squares regression) と呼ばれている。なお、この問題はパラメタ $\boldsymbol{\xi}$ を用いるよりもモデルの最小2乗予測値 \hat{v} を用いて説明した方がわかりやすいので、しばらくの間、パラメタ $\boldsymbol{\xi}$ を無視してモデルの最小2乗推定値 \hat{v} を用いて議論を進める。すなわち、

$$\text{RSS} = \sum_{i=1}^N (v_i - \hat{v}_i)^2$$

を最小とするような \hat{v} を

$$u_i < u_{i'} \rightarrow \hat{v}_i \leq \hat{v}_{i'} \quad (25)$$

$$u_i = u_{i'} \rightarrow \hat{v}_i = \hat{v}_{i'} \quad (26)$$

という制約条件の下で求めることを考える。この解は、 v と u の順序が等しければ、 $\hat{v} = v$ となることは明かであるが、もしも v と u の順序が全く逆転していたら、 $\hat{v} = \bar{v}$ となる。したがって、求める \hat{v} は一般にこの両者の中間となる。

なお、通常の回帰分析同様、 \hat{v} はあくまで説明変数 u の関数 (単調増加) であるが、最小2乗単調回帰では、 \hat{v} は v の部分の平均として計算される。

このような \hat{v} を求めるためには、まず、 v と u を u の小さい順に並べ替え、同順位の u に対応する v の要素の平均を取ったものを \hat{v} とする。その後、 \hat{v} を小さい方から調べていき、順序の逆転が起これば逆転が無くなるまで v の小さな方向に向かって平均を取るという作業を行う。

以下に最小2乗単調回帰を行うためのアルゴリズムを示す。なお、このアルゴリズムは Kruskal(1964b)の方法を改善したものである。

```

最小2乗単調回帰
 $\hat{v} = v$ ; ただし、同順位の  $u$  に対応する  $v$  は平均する。;
do i=2 to N;
  do j=1 to i-1;
    if  $\hat{v}_{i-j} \leq \hat{v}_i$  then do;
       $\hat{v}_{i-j}$  から  $\hat{v}_i$  までをその平均値  $\sum_{k=i-j}^i \hat{v}_k / (j+1)$  で置き換える。
    end;
  end;
end;
end;

```

上記のアルゴリズムを表 1 の第 2 列と 3 列に示されたデータに適用した結果を同表の第 6 から 14 列に示す。第 4 列に同順位の平均を取った v の値が示されている。第 14 列に示された値が RSS を最小とする \hat{v} の値である。なお、上記のアルゴリズムの中で、if 文の条件が偽になった場合はなにもしないため表から除いてある。また、 u と v および \hat{v} をプロットしたものを図 2 に示す。図中、* で示したものがデータで、M が \hat{v} を表している。なお、残差の 2 乗和は RSS=19 となり、通常の回帰分析の残差の 2 乗和 RSS(linear) = 37.91 よりも小さくなっている。また、分散説明率は、 $R^2=0.8177$ であった。

なお、最小 2 乗単調回帰は単調性の条件 (25) 式と (26) 式を

$$u_i < u_{i'} \rightarrow \hat{v}_i \leq \hat{v}_{i'} \quad (27)$$

と置き換えても定義することが出来る。(この条件は、たとえば SAS の中 proc mds では untie 条件、先に述べたものは monotone 条件と呼ばれている。また、PROC IML には monotone 条件の下で解を求める関数 opscal が用意されている。) この場合は上記のアルゴリズムの中で、同順位の u に対する v の平均で \hat{v} を定義したところを変更し、同順位の u に対する v を v の順序で並べ替えたものを \hat{v} とすればよい。上記のアルゴリズムを表 1 の第 2 列と 3 列に示されたデータに適用した結果を表 2 の第 6 列以降に示す。第 4 列に同順位の部分を並べ替えた v の値が示されている。また、第 8 列をもとの v の値に並べ替えたものが第 9 列であり、それが RSS を最小とする \hat{v} の値である。なお、上記のアルゴリズムの中で、if 文の条件が偽になった場合はなにもしないため表から除いてある。また、 u と v および \hat{v} をプロットしたものを図 3 に示す。図中、* で示したものがデータで、M が \hat{v} を表している。なお、残差の 2 乗和は RSS=13.5 となり、同順位の制約を入れたものよりも小さくなっている。また、分散説明率は、 $R^2=0.8705$ であった。

この最小 2 乗単調回帰は通常の線形モデルと同様に、展開行列 M を用いて

$$\theta(u | \xi) = M(u)\xi$$

という形で記述することが出来る。ただし、 $M(u)$ は上記のアルゴリズムで求めた \hat{v} を

$$\hat{\xi} = (M(u)'M(u))^{-1}M(u)'v \quad \text{and} \quad \hat{v} = M(u)\hat{\xi} \quad (28)$$

という形で求めることを可能にするような行列で、たとえば、表 1 の例では

$$M(\mathbf{u}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \hat{\xi} = \begin{bmatrix} 2.0 \\ 4.5 \\ 9.0 \\ 12.0 \end{bmatrix}$$

となる。ただ、この $M(\mathbf{u})$ 行列は、実際はどの部分に 1 を置くかということに関して \mathbf{v} に依存しており、RSS を最小とするように推定されたモデルのパラメタの一部であることに注意しておく必要がある。また、この最小 2 乗基準単調回帰のための M 行列は常に $\mathbf{1}$ をその列空間に含むことになるが、このことは基準化された最小 2 乗基準を最小化する際に重要となる。

最後に、単調性の条件をはずし、単に等号条件のみにした条件の下での最小 2 乗解 \hat{v} は、 \mathbf{u} の同順位のもの \mathbf{v} の平均となるが、この解は Fisher の最適変換、もしくは名義尺度の数量化と呼ばれることがある。なお、この場合の最適変換は \mathbf{u} の同順位の位置にのみ依存するので、単なる変数の展開として取り扱える。

表 1: 最小 2 乗単調回帰の説明 (monotone)

obs	\mathbf{u}	\mathbf{v}	tie	添え字									
				i	3	4	4	5	5	5	7	8	8
				j	1	1	2	1	2	3	1	1	2
1	1	2	2		2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
2	2	5	5		4.75	4.75	4.66	4.66	4.66	4.50	4.50	4.50	4.50
3	4	7	4.5		4.75	4.62	4.66	4.66	4.44	4.50	4.50	4.50	4.50
4	4	2	4.5		4.50	4.62	4.66	4.33	4.44	4.50	4.50	4.50	4.50
5	5	4	4		4.00	4.00	4.00	4.33	4.44	4.50	4.50	4.50	4.50
6	6	10	10		10.00	10.00	10.00	10.00	10.00	10.00	9.25	9.25	9.00
7	8	10	8.5		8.50	8.50	8.50	8.50	8.50	8.50	9.25	8.87	9.00
8	8	7	8.5		8.50	8.50	8.50	8.50	8.50	8.50	8.50	8.87	9.00
9	9	12	12		12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00

参考文献

- 前川眞一 (1997) SASによる多変量データの解析 竹内啓監修 SASで学ぶ統計的データ解析
4. 東京大学出版会
- van Wezel, M.C. & Kusters, W.A. Nonmetric Multidimensional Scaling: Neural Networks versus Traditional Techniques. *Intelligent Data Analysis 8 (2004) 601 - 613. IOS Press*
<http://people.few.eur.nl/mvanwezel/papers/ida00191.pdf>

表 2: 最小2乗単調回帰の説明 (untie)

obs	u	v	sorted	添え字				sorted
				i	3	5	7	
				j	1	1	1	
1	1	2	2		2.0	2.0	2.0	2.0
2	2	5	5		3.5	3.5	3.5	3.5
3	4	7	2		3.5	3.5	3.5	5.5
4	4	2	7		7.0	5.5	5.5	3.5
5	5	4	4		4.0	5.5	5.5	5.5
6	6	10	10		10.0	10.0	8.5	8.5
7	8	10	7		7.0	7.0	8.5	10.0
8	8	7	10		10.0	10.0	10.0	8.5
9	9	12	12		12.0	12.0	12.0	12.0

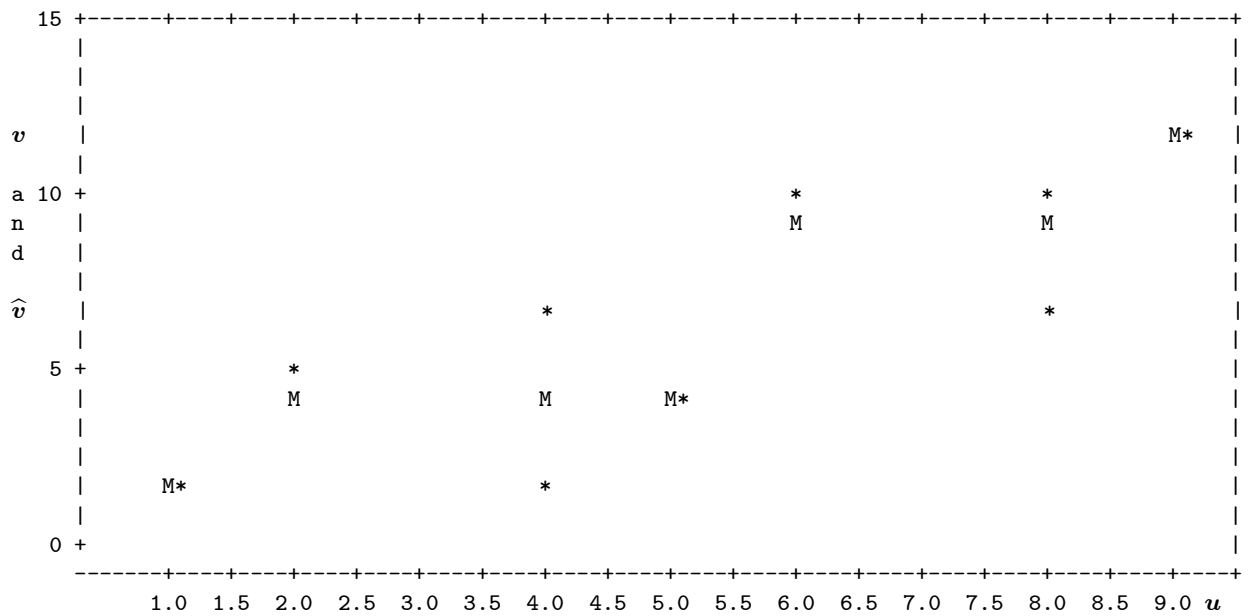


図 2: 最小2乗単調回帰 (monotone)

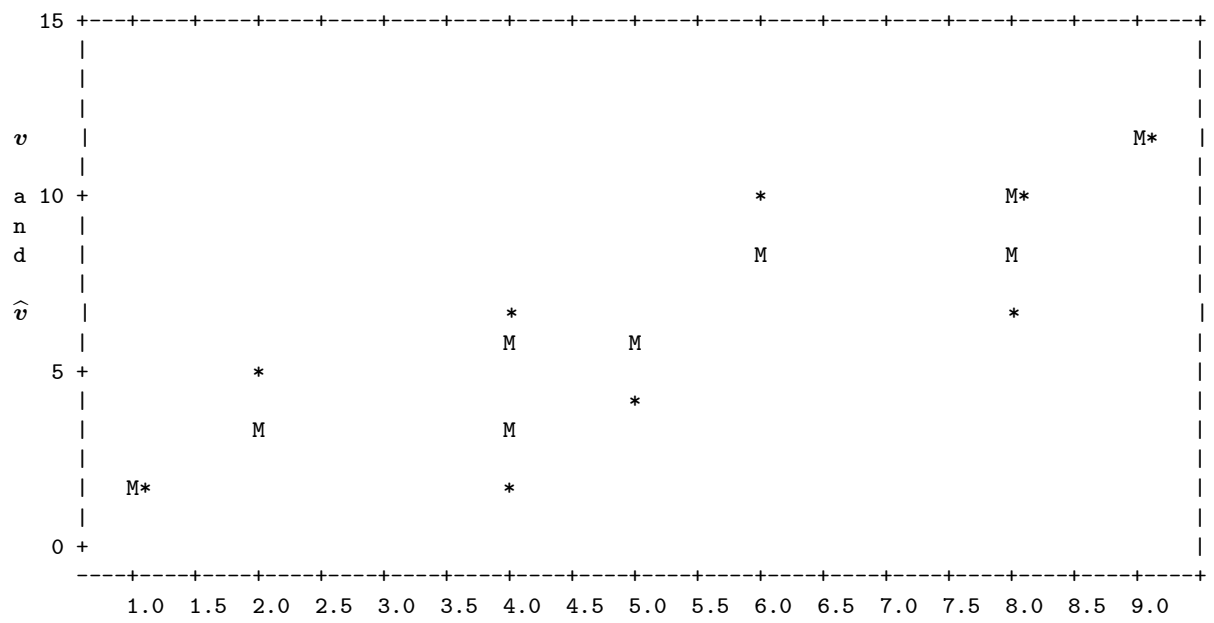


图 3: 最小二乘单调回归 (untie)