線形回帰と最小二乗法

回帰と分類
線形回帰
最小二乗法

予測の問題：入力と出力

予測モデル

入力：説明変数(explanatory variable)
独立変数(independent variable)
予測変数(predictor)

$$y = f(x; \Theta)$$

出力：目的変数(objective variable)
従属変数(dependent variable)
応答変数(response variable)

回帰と分類

予測モデル

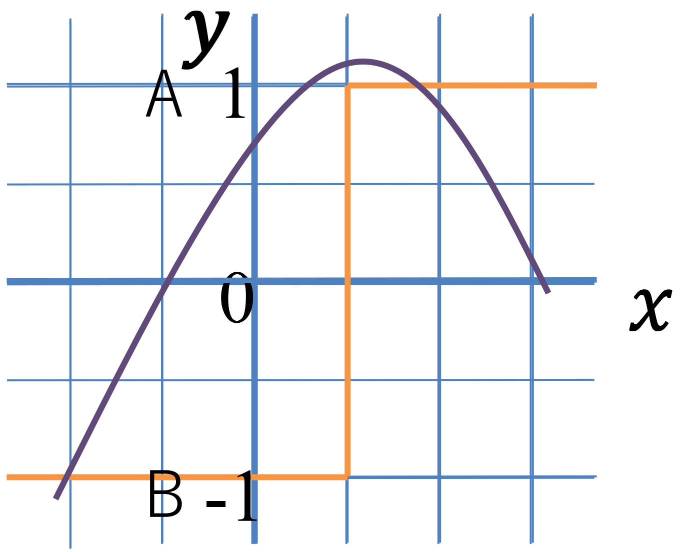
入力：

$$y = f(x; \Theta)$$

出力：

y が連続値のとき 回帰

y が離散値のとき 分類／識別



回帰／識別関数とパラメタ

予測モデル

入力:

回帰関数／識別関数
回帰モデル／識別モデル

$$y = f(x; \Theta)$$

パラメタ

出力:

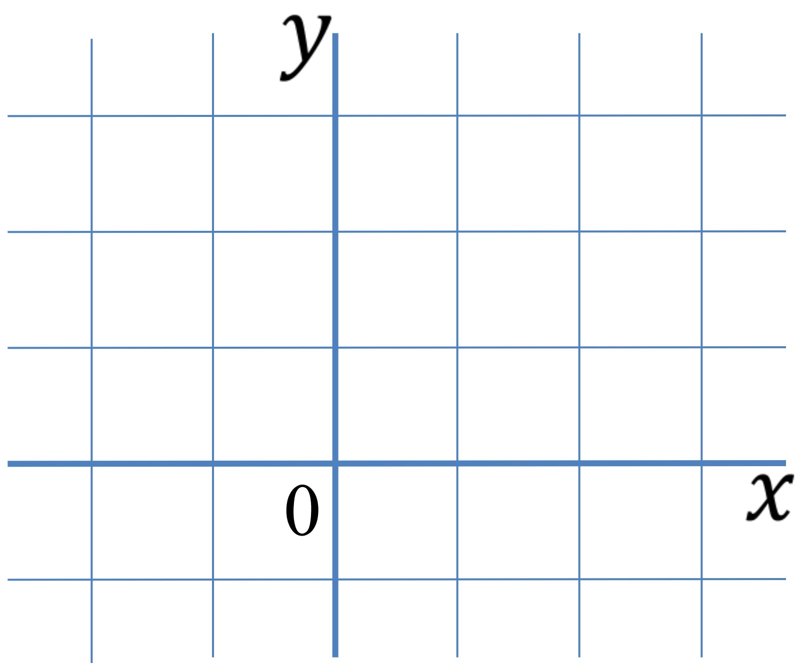
y が連続値のとき 回帰
y が離散値のとき 分類／識別

x と y とを関係付ける関数の形は，パラメタ θ で特徴づけられる。

回帰関数のパラメタ

$$y = f(x; \Theta)$$

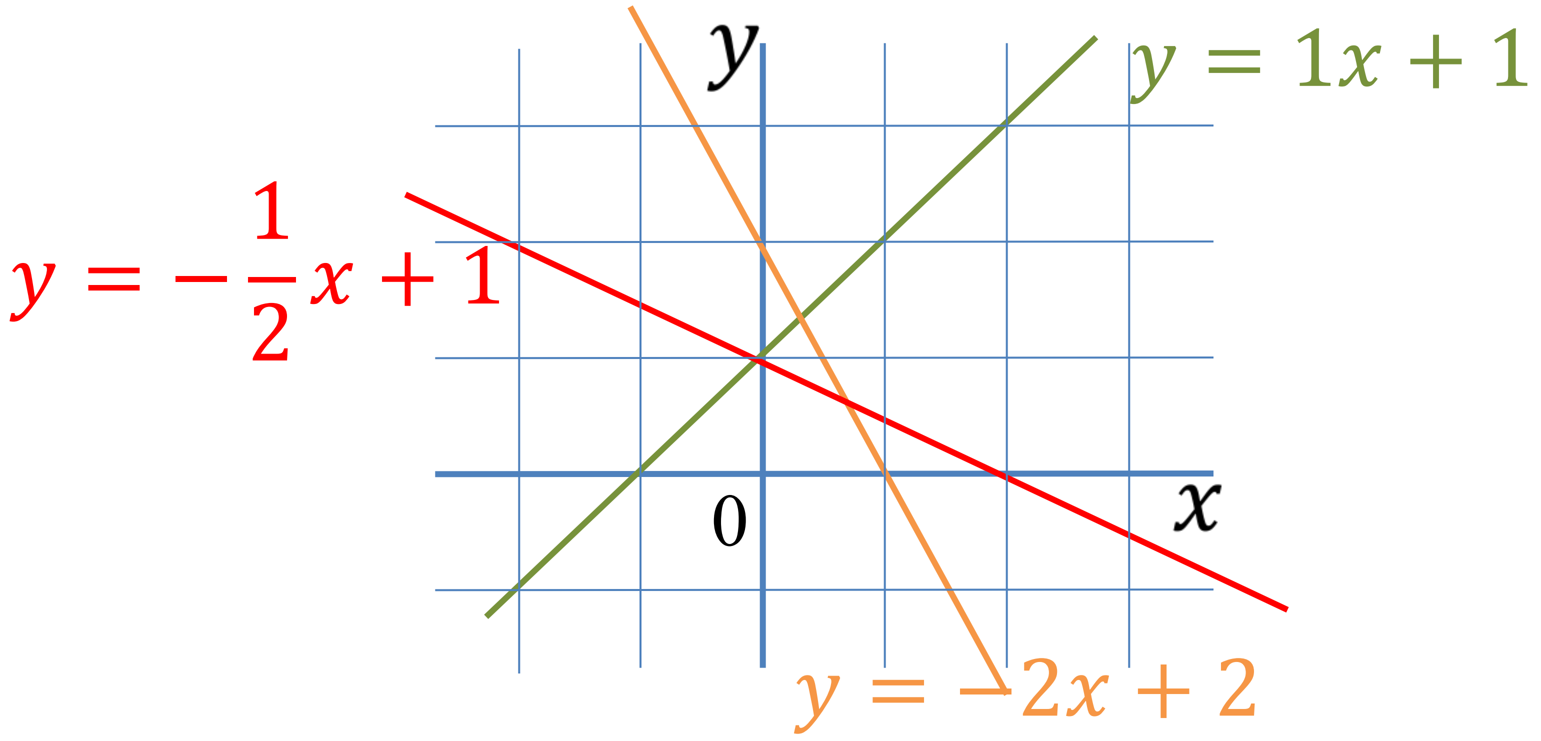
$$y = ax + b, \quad \Theta = (a, b)$$



回帰関数のパラメタ

$$y = f(x; \Theta)$$

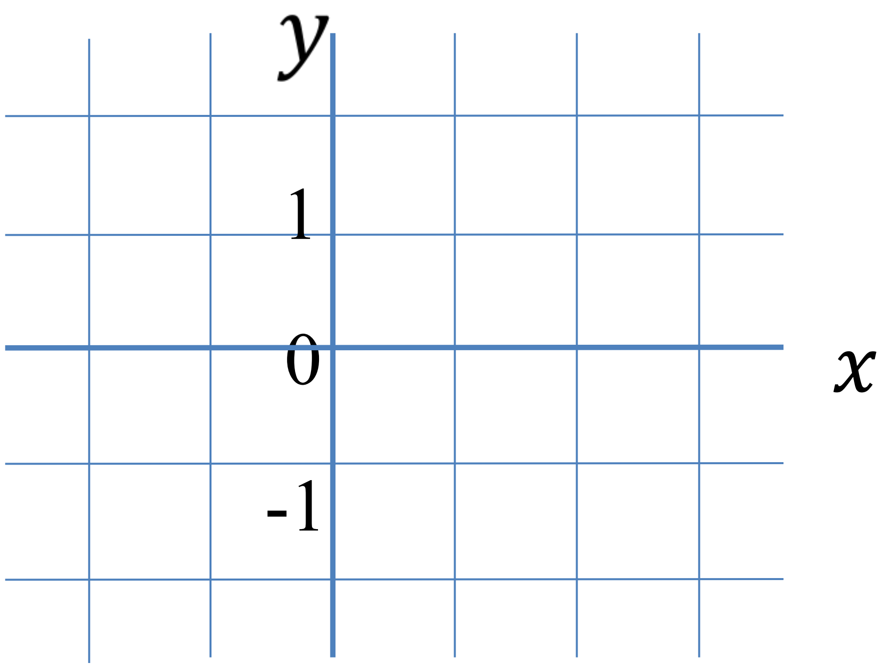
$$y = ax + b, \quad \Theta = (a, b)$$



識別関数のパラメタ

$$y = f(x; \Theta)$$

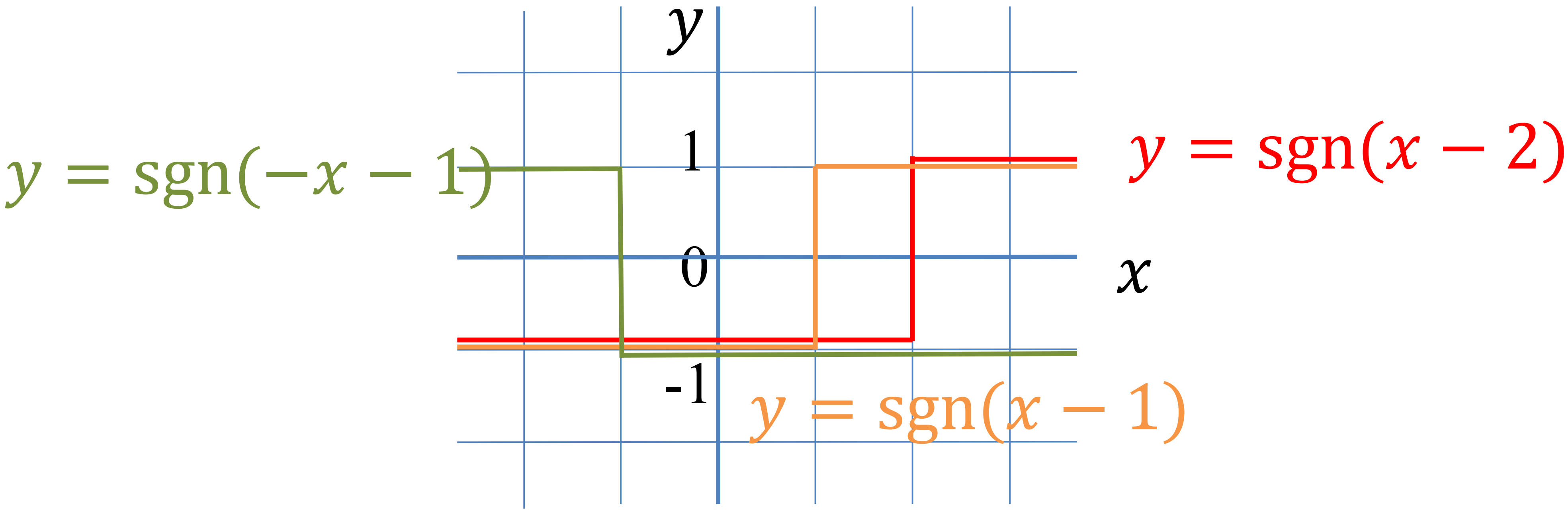
$$y = sgn(ax + b),\ \Theta = (a, b)$$



識別関数のパラメタ

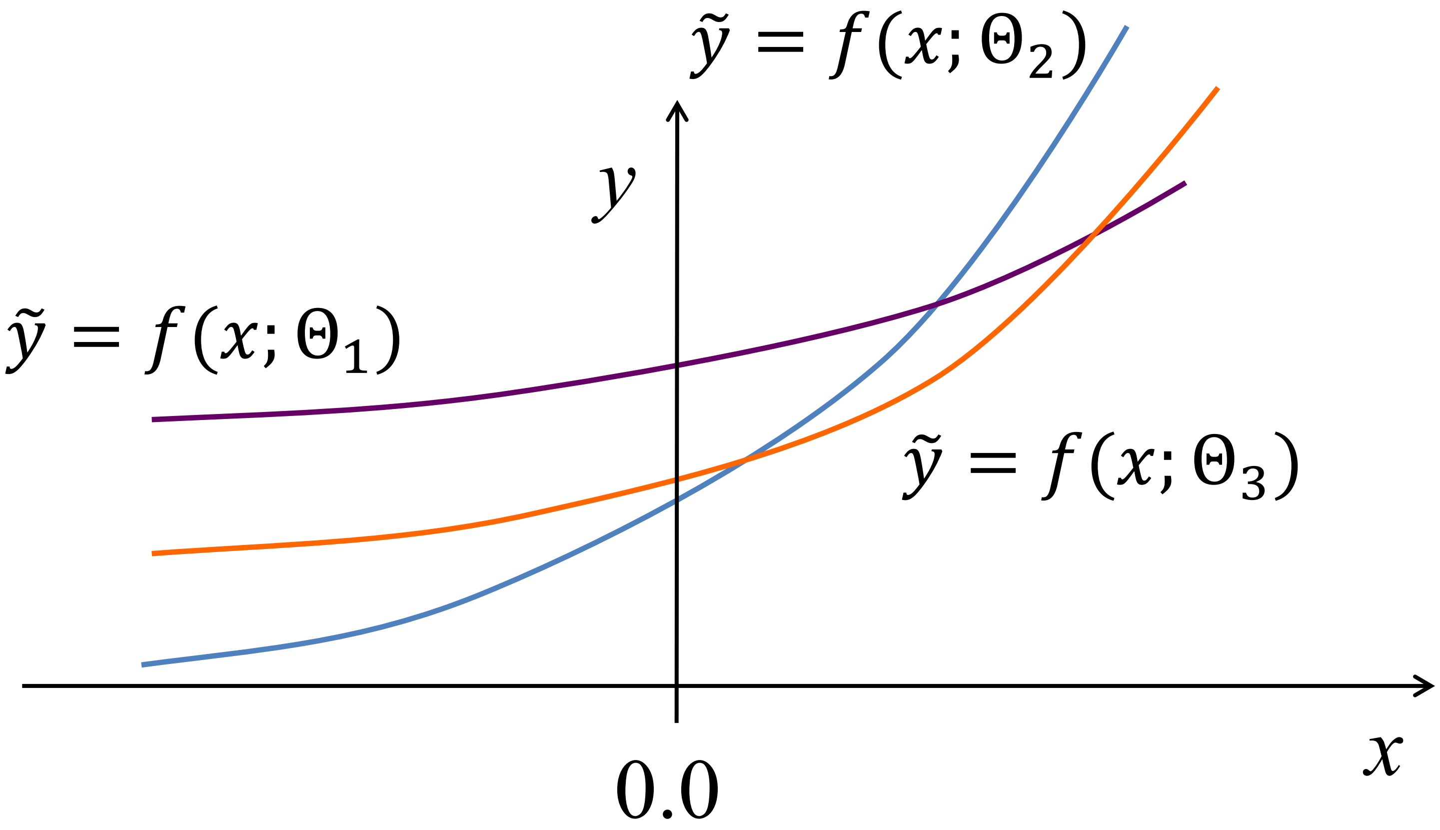
$$y = f(x; \Theta)$$

$$y = sgn(ax + b), \Theta = (a, b)$$



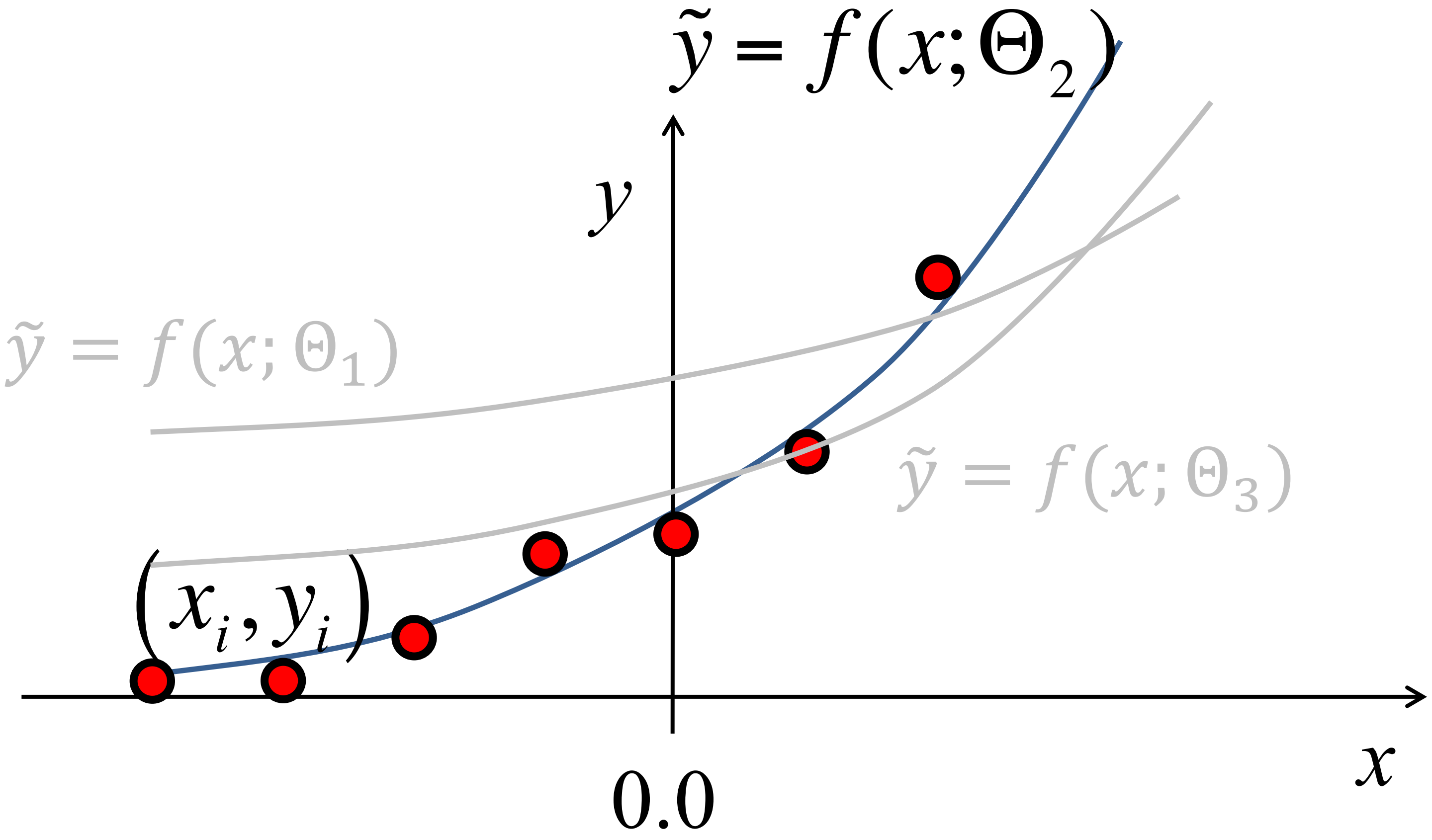
教師つき学習

パラメタ Θ が異なると，様々な予測モデルができる。



教師つき学習

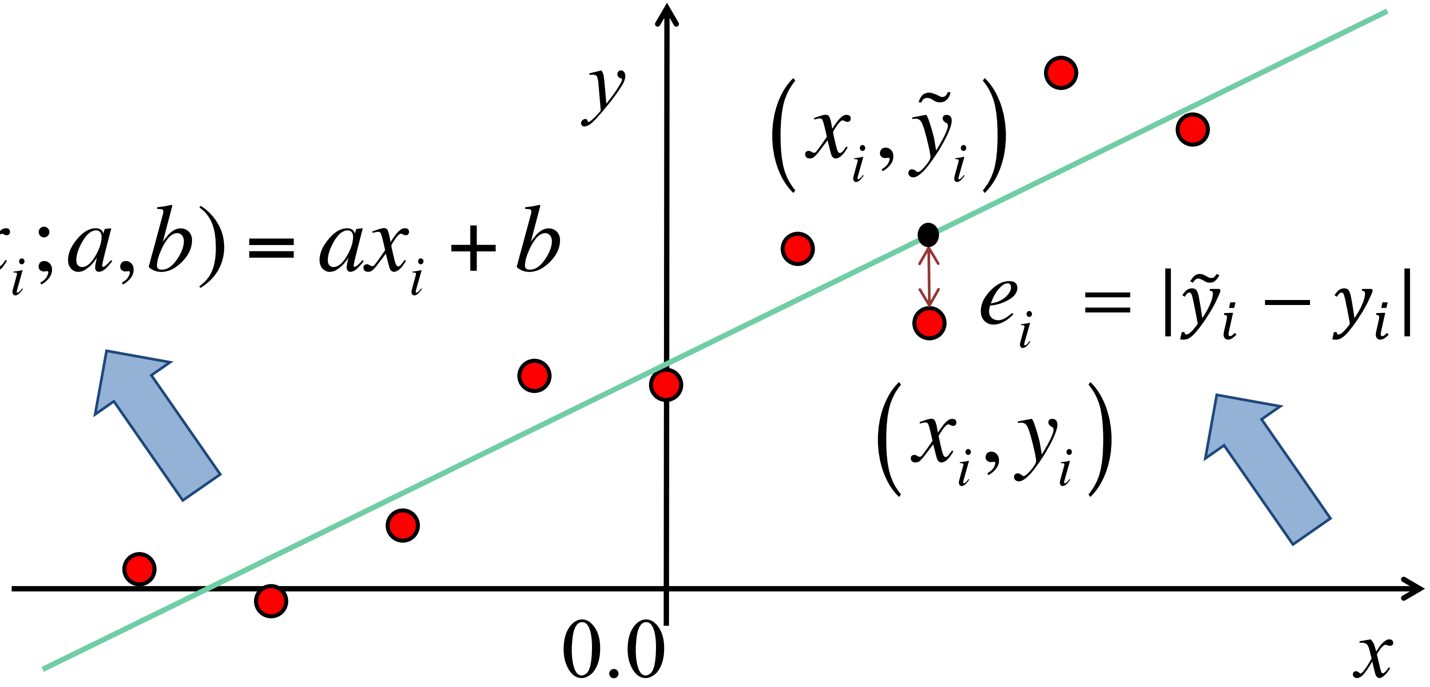
モデルのパラメタ Θ を，学習データ(training data)を用いて推定
⇒　教師あり学習



線形回帰

- 与えられたデータに対し，線形関数を当てはめる問題

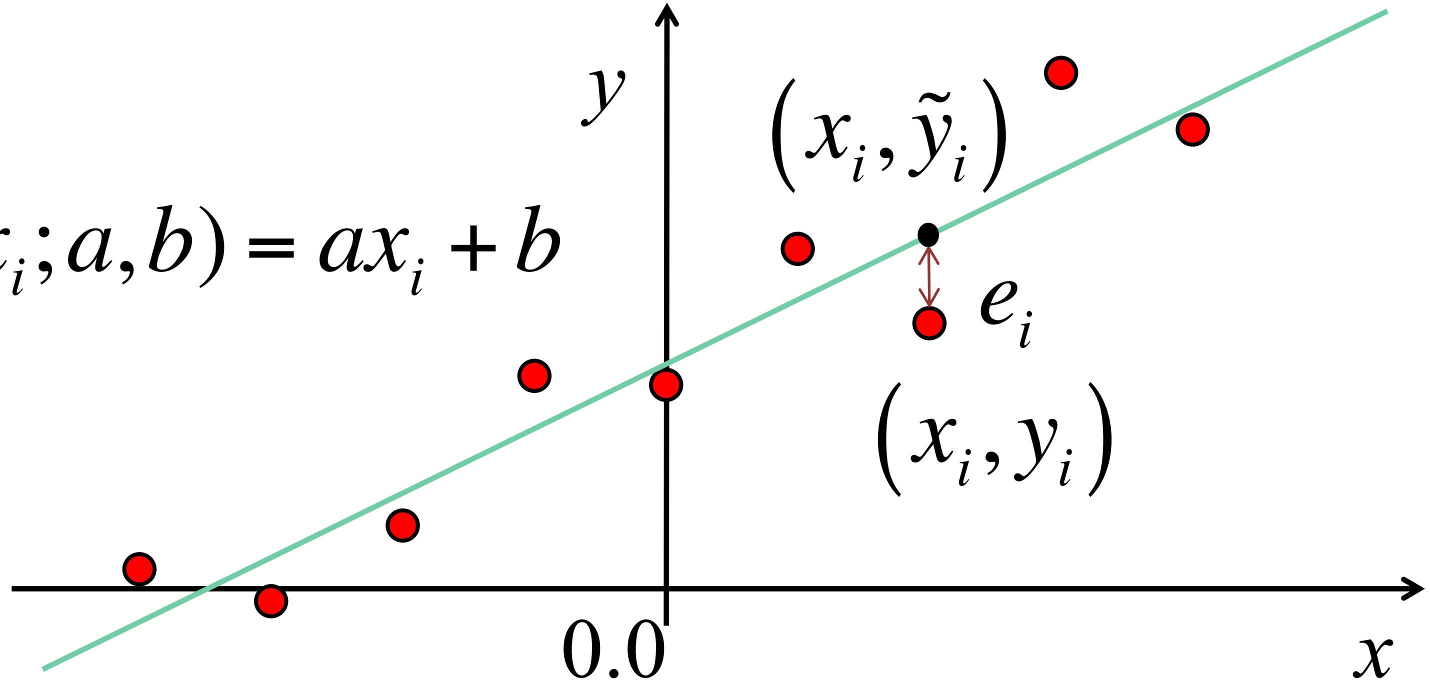
$$\tilde{y}_i = f(x_i;a,b) = ax_i + b$$



線形回帰

- 与えられたデータに対し，線形関数を当てはめる問題

$$\tilde{y}_i = f(x_i; a, b) = a x_i + b$$



二乗誤差の最小化： $\min_{a,b} \sum_{i=1}^{N} e_i^2 = \min_{a,b} \sum_{i=1}^{N} (\tilde{y}_i - y_i)^2$

最小二乗法

n 番目の学習データを (x_n, y_n) とする。

x_n に要素 1 を付け加えてベクトル化したものを $\boldsymbol{x}_n$ とし，
回帰直線の係数ベクトルを $\boldsymbol{w}$ とする。

$$\boldsymbol{x}_n = \begin{pmatrix} x_n \\ 1 \end{pmatrix}, \quad \boldsymbol{w} = \begin{pmatrix} a \\ b \end{pmatrix}$$

寄り道：ベクトルの1次元拡張

　N 次元ベクトル $\boldsymbol{x} = (x_1\ x_2\ \cdots\ x_N)^T$を扱うとき，定数項が生じる場合がある。

　例えば，超平面の方程式は，その法線ベクトル $\boldsymbol{w} = (w_1\ w_2\ \cdots\ w_N)^T, \|\boldsymbol{w}\| = 1$，原点との距離 c によって，

$$\boldsymbol{w}^T\boldsymbol{x} = c$$

と表される。

　すなわち，平面の方程式は，$\boldsymbol{w}$，c という２種類のパラメタを必要とする。

　このため，パラメタを推定する場合などは，この $\boldsymbol{w}$，c のそれぞれについて異なる扱いが必要となる。

寄り道：ベクトルの1次元拡張

　しかし，ここで，

$$\boldsymbol{x} = (x_1\ x_2\ \cdots\ x_N\ x_{N+1})^T,\ x_{N+1} = 1$$

$$\boldsymbol{w} = (w_1\ w_2\ \cdots\ w_N\ w_{N+1})^T,\ w_{N+1} = -c$$

とおくと，方程式は，

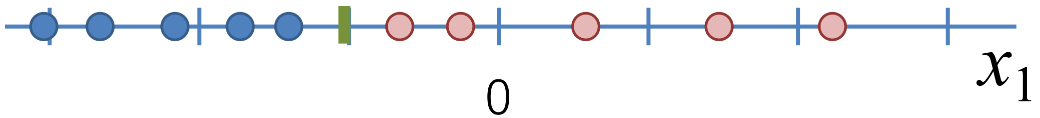
$$\boldsymbol{w}^T\boldsymbol{x} = 0$$

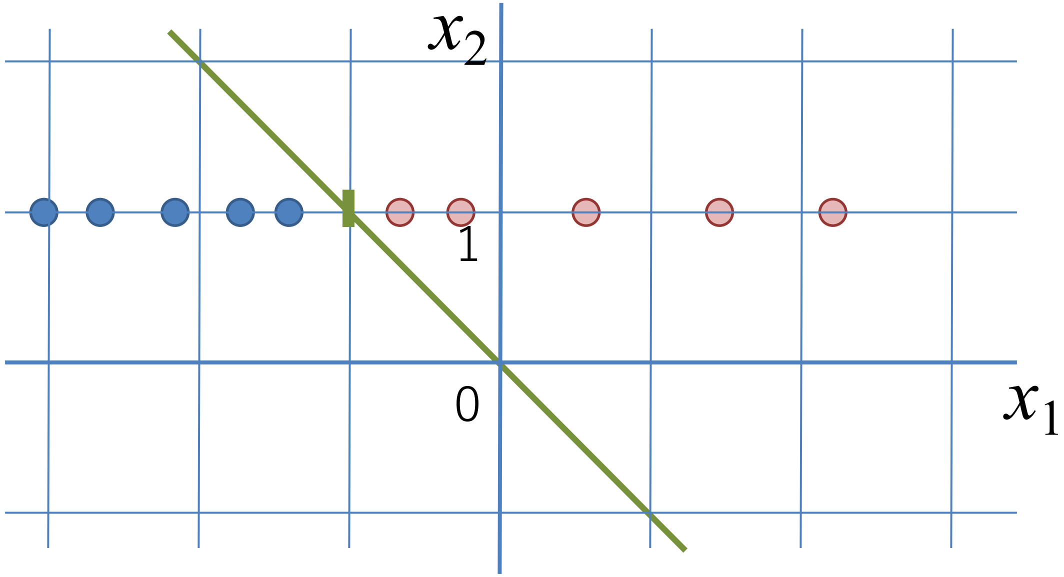
となって，形式上 c を除くことができる。

　この「定数項が現れない」という性質は，アルゴリズム記述上便利なので，ベクトルに 常に 1 となる要素を付加する操作を行うことが多い。

寄り道： ベクトルの1次元拡張

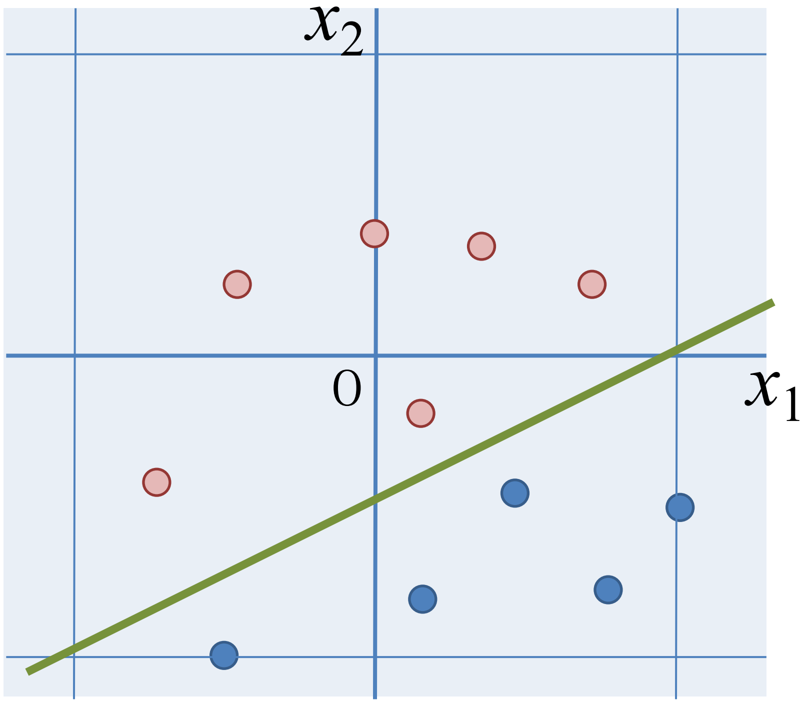
1次元データの2次元拡張の例

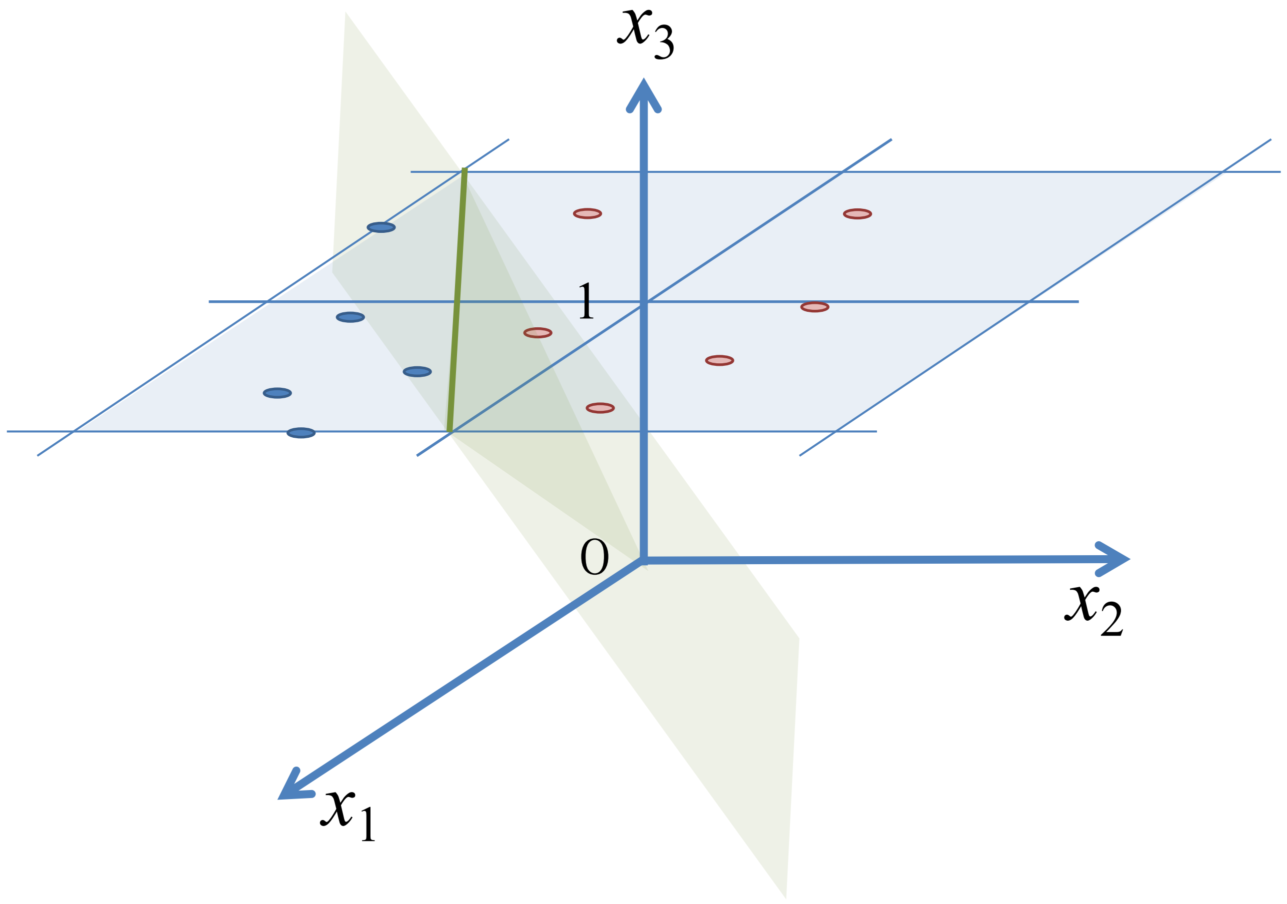




寄り道： ベクトルの1次元拡張

2次元データの3次元拡張の例





最小二乗法

再掲

n 番目の学習データを (x_n, y_n) とする。

x_n に要素 1 を付け加えてベクトル化したものを $\boldsymbol{x}_n$ とし，
回帰直線の係数ベクトルを $\boldsymbol{w}$ とする。

$$\boldsymbol{x}_n = \begin{pmatrix} x_n \\ 1 \end{pmatrix}, \quad \boldsymbol{w} = \begin{pmatrix} a \\ b \end{pmatrix}$$

最小二乗法

n 番目の学習データを (x_n, y_n) とする。

x_n に要素 1 を付け加えてベクトル化したものを $\boldsymbol{x}_n$ とし，
回帰直線の係数ベクトルを $\boldsymbol{w}$ とする。

$$\boldsymbol{x}_n = \begin{pmatrix} x_n \\ 1 \end{pmatrix}, \quad \boldsymbol{w} = \begin{pmatrix} a \\ b \end{pmatrix}$$

このとき，n番目の予測値を $\tilde{y}_n$ とすると，

$$\tilde{y}_n = a x_n + b = {\boldsymbol{x}_n}^T \boldsymbol{w}$$

と書ける。

n 番目の行ベクトルとして $\boldsymbol{x}_n{}^T$ を持つ行列を D とする。

$$D = \begin{pmatrix} \boldsymbol{x}_1{}^T \\ \vdots \\ \boldsymbol{x}_n{}^T \\ \vdots \\ \boldsymbol{x}_N{}^T \end{pmatrix}$$

このとき，推定値のベクトルは，

$$\tilde{\boldsymbol{y}} = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_n \\ \vdots \\ \tilde{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1{}^T \\ \vdots \\ \boldsymbol{x}_n{}^T \\ \vdots \\ \boldsymbol{x}_N{}^T \end{pmatrix} \boldsymbol{w} = D\boldsymbol{w}$$

誤差ベクトルは，

$$\boldsymbol{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \\ \vdots \\ e_N \end{pmatrix} = \begin{pmatrix} \tilde{y}_1 - y_1 \\ \vdots \\ \tilde{y}_n - y_n \\ \vdots \\ \tilde{y}_N - y_N \end{pmatrix} = \tilde{\boldsymbol{y}} - \boldsymbol{y} = D\boldsymbol{w} - \boldsymbol{y}$$

このとき，二乗誤差 E は，

$$E = \frac{1}{2}\sum_{n=1}^{N} {e_n}^2 = \frac{1}{2}\boldsymbol{e}^T\boldsymbol{e} = \frac{1}{2}(D\boldsymbol{w} - \boldsymbol{y})^T(D\boldsymbol{w} - \boldsymbol{y})$$

$$= \frac{1}{2}(\boldsymbol{w}^T D^T D\boldsymbol{w} - y^T D\boldsymbol{w} - \boldsymbol{w}^T D^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{y})$$

$$E = \frac{1}{2}(\boldsymbol{w}^T D^T D\boldsymbol{w} - y^T D\boldsymbol{w} - \boldsymbol{w}^T D^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{y})$$

これを最小化する$\boldsymbol{w}$ を求めるために，傾きを求める。

$$\frac{\partial E}{\partial \boldsymbol{w}} = \frac{1}{2}(D^T D\boldsymbol{w} + (\boldsymbol{w}^T D^T D)^T - (y^T D)^T - D^T \boldsymbol{y})$$

$$= D^T D\boldsymbol{w} - D^T \boldsymbol{y}$$

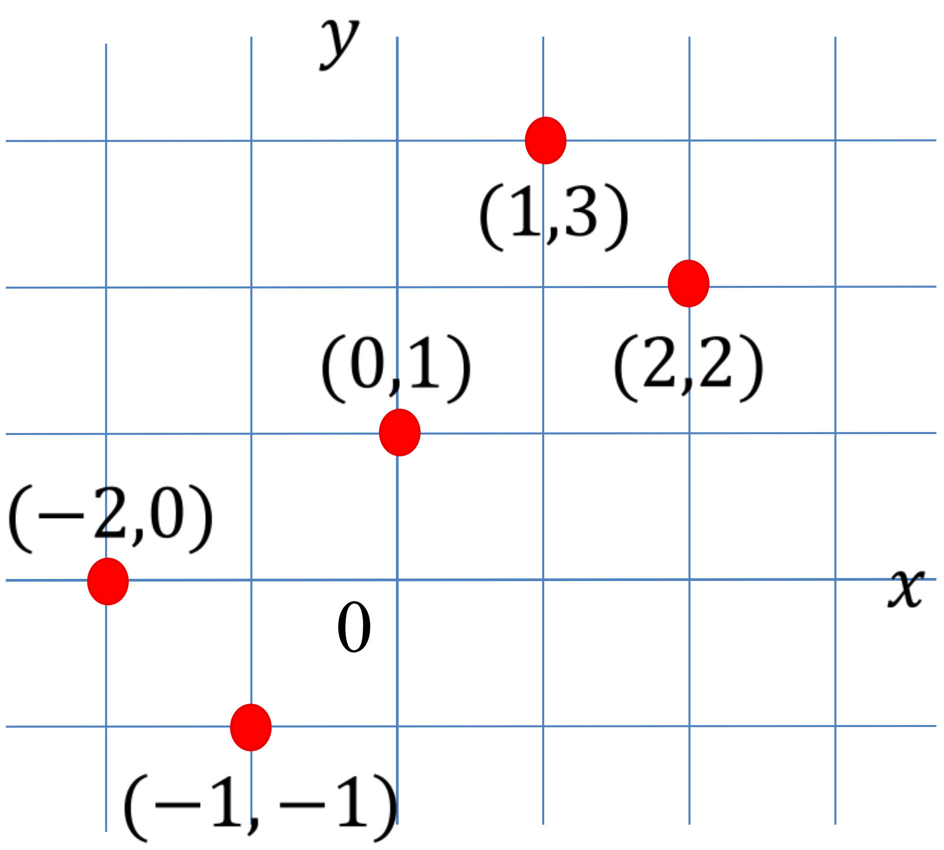
これを0とおいて，$\boldsymbol{w}$ について解くことで，

$$\boldsymbol{w} = (D^T D)^{-1} D^T \boldsymbol{y}$$

と求まる

例題

次の５点から，線形回帰モデルを，最小二乗基準で求めよ。

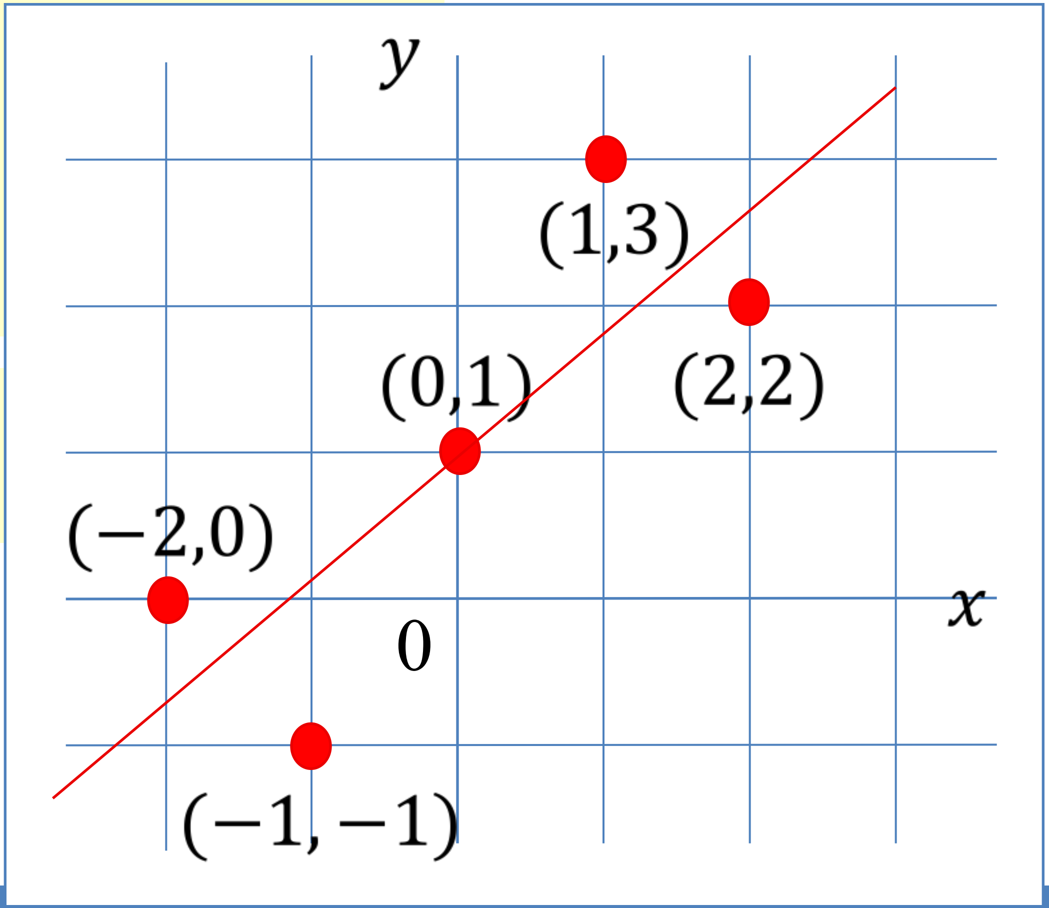


```
import numpy as np
from scipy import linalg as la

D = np.matrix([[-2,1],[-1,1],[0,1],[1,1],[2,1]])
y = np.matrix([0,-1,1,3,2])
y = y.T
A = D.T * D
w = la.inv(A) *D.T*y

print w
```

```
[[0.8]
[1.0]]
```



説明変数が多変数の場合

$(x_{n1}, x_{n2}, \cdots, x_{np}, y_n)$: n番目の学習データ

$$\boldsymbol{x}_n = [x_{n1}, x_{n2}, \cdots, x_{np}, 1]^T$$

$D = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_i, \cdots, \boldsymbol{x}_N]^T$: データ行列

$\boldsymbol{w} = [a_1, a_2, \cdots, a_p, b]^T$: パラメタ

$\tilde{\boldsymbol{y}} = D\boldsymbol{w}$: 予測式

$$\boldsymbol{w} = (D^T D)^{-1} D^T y$$

まとめ

- 入力 $\boldsymbol{x}$ から出力 y を予測する問題において，
  y が連続値の場合を 回帰 とよび，
  y が離散値のとき 分類／識別 と呼ぶ。
- y を $\boldsymbol{x}$ の線形関数で表す問題を線形回帰と呼ぶ。
- 二乗誤差の最小化を基準として予測関数（回帰関数／識別関数）を求める方法を最小二乗法と呼ぶ。
- 線形回帰の問題を最小二乗法で解くとき，解は解析的に求まる。

演習問題

1. 次の12点から，線形回帰モデルを，最小二乗基準で求めよ。

$$(0,0,1),(0,1,3),(0,2,2),(0,3,3)$$
$$(1,0,1),(1,1,4),(1,2,5),(1,3,4)$$
$$(2,0,4),(2,1,5),(2,2,4),(2,3,7)$$