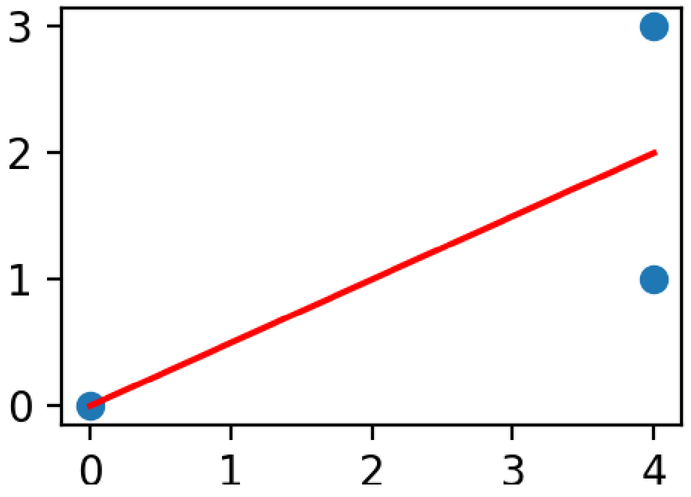
問題：最小二乗法により、３点（0, 0），（4, 3），（4, 1）に当てはまる直線を求めよ。




```
import numpy as np;  import matplotlib.pyplot as plt

# 人工データ３点をプロット
x = [0, 4, 4]
y = [0, 3, 1]
plt.scatter(x, y)

#回帰直線の傾きと切片
c = np.polyfit(x,y,1);   plt.plot([0,4], [c[1],4*c[0]+c[1]], color="red")
# 傾き c[0]              切片 c[1]
# [5.00000000e-01, 4.97237083e-16]
```

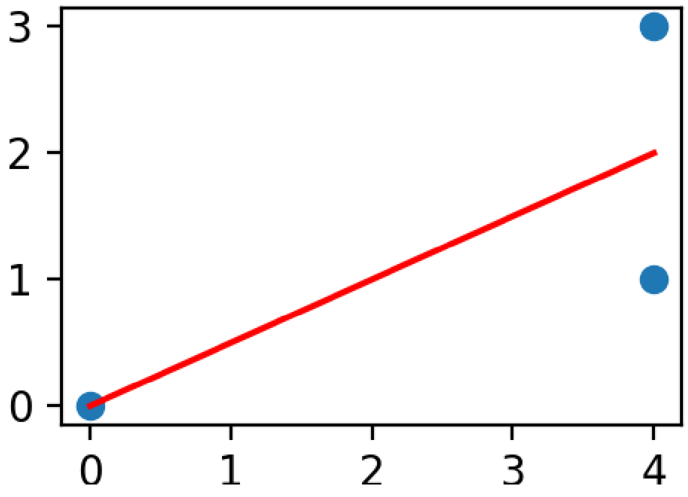
[解答] $\hat{y} = ax + b$ の形に回帰直線を仮定する。

$$Q := (b-0)^2 + (4a+b-3)^2 + (4a+b-1)^2$$

← この2乗誤差Qを最小にするような (a, b) を求める！

$$\begin{cases} 0 = \dfrac{\partial Q}{\partial a} = 4a + b - 2 \\ 0 = \dfrac{\partial Q}{\partial b} = 8a + 3b - 4 \end{cases} \Rightarrow \begin{cases} a = 0.5 \\ b = 0 \end{cases}$$

問題：最小二乗法により、３点（0, 0），（4, 3），（4, 1）に当てはまる直線を求めよ。




```
import numpy as np;  import matplotlib.pyplot as plt

# 人工データ３点をプロット
x = [0, 4, 4]
y = [0, 3, 1]
plt.scatter(x, y)

#回帰直線の傾きと切片
c = np.polyfit(x,y,1);   plt.plot([0,4], [c[1],4*c[0]+c[1]], color="red")
# 傾き c[0]              切片 c[1]
# [5.00000000e-01, 4.97237083e-16]
```

[解答]　$\hat{y} = ax + b$ の形に回帰直線を仮定する。

$$Q := (b-0)^2 + (4a+b-3)^2 + (4a+b-1)^2$$

← この2乗誤差Qを最小にするような (a, b) を求める！

$$\begin{cases} 0 = \dfrac{\partial Q}{\partial a} = 4a + b - 2 \\ 0 = \dfrac{\partial Q}{\partial b} = 8a + 3b - 4 \end{cases} \Rightarrow \begin{cases} a = 0.5 \\ b = 0 \end{cases}$$

一般論：傾きβと切片αはx_iとy_iで表せる

$$Q := \sum_{i=1}^{n} (\underbrace{ax_i + b}_{\text{直線の予測値}} - \underbrace{y_i}_{\text{観測値}})^2$$

を最小化することで、次のaとbが得られる：

$$\begin{cases} a = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ b = \underbrace{\bar{y}}_{\text{平均値}} - a\bar{x} \end{cases}$$

← **xとyの共分散**

導出の式変形までも理解したい人は
「数理統計学」鈴木・山田（1996）
などを参照のこと。

$Q := \sum_{i=1}^{n}(ax_i + b - y_i)^2$