In [1]:
```
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
from sklearn import linear_model
```

1,(a) y = beta0 + beta1*x1* + *beta2*x2

In [13]:
```
X = np.array([[0,1,0,1],[0,0,1,1]]).T
y = np.array([2,4,4,5])
```

In [17]:
```
nsamp = len(y)
ones = np.ones((nsamp,1))
A = np.hstack((ones,X))
out = np.linalg.lstsq(A,y)
beta = out[0]
beta
```

```
/opt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4: FutureWarning: `rcond` parameter will change to t
he default of machine precision times ``max(M, N)`` where M and N are the input matrix dimensions.
To use the future default and silence this warning we advise to pass `rcond=None`, to keep using the old, explicitly
pass `rcond=-1`.
  after removing the cwd from sys.path.
```

Out[17]:
```
array([2.25, 1.5 , 1.5 ])
```

1. (b) beta0 = 2.25 beta1 = 1.5 beta2 = 1.5

In [15]:
```
# Create linear regression object
regr = linear_model.LinearRegression()
regr.fit(X,y)
y_pred = regr.predict(X)
RSS = np.sum((y_pred-y)**2)
Rsq = 1-RSS/len(y)/(np.std(y)**2)
Rsq
```

Out[15]:
```
0.9473684210526316
```

1,(c) R^2 is 0.9474

2. (a) target variable y is the age at which the clients died.

(b) let blood pressure as x_1, heart rate as x_2, BMI as x_3.
The modell is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

(c) Use on-hot encoding to encode the gender
use non-redaundant on-hot encoding. let gender as x_4, ~~is~~ when the client is female $x_4 = 1$ when the client is male. $x_4 = 0$

the model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4.$$

Reason: one-hot coding allows a different intercept for each catogory. The redundant on-hot coding can avoid the problem of making A^TA non-invertible.

3. (a) $$y_k \approx \sum_{\ell=1}^{L} a_\ell \cos(\Omega_\ell k) + b_\ell \sin(\Omega_\ell k)$$

$$\underline{y} \approx \underline{A}\underline{\beta}$$

$$\underline{y} = [y_0\, y_1\, y_2 \cdots y_{N-1}]$$

$$\underline{\beta} = [\beta_0\, \beta_1 \cdots \beta_{2L}] = [\beta_0\, a_1\, a_2 \cdots a_L\, b_1\, b_2 \cdots b_L]$$

$$\underline{A} = [\underline{1}\ \ \underline{x}]$$

~~$\underline{x} = [\cos(\Omega_1 \cdot 0)\quad \cos(\Omega_2 \cdot 0) \cdots \cos(\Omega_L \cdot 0) \cdot \sin(\Omega_1 \cdot 0) \cdots$~~
~~$\cos(\Omega_1 \cdot 1)\quad \cos(\Omega_2 \cdot 1) \cdots \cos(\Omega_L \cdot 1) \cdot \sin(\Omega$~~
~~$\cos(\Omega_1 \cdot k)$~~
$\vdots$

3.(a) continue:

$$\underline{X} = \begin{bmatrix} \cos(\Omega_1 \cdot 0) & \cos(\Omega_2 \cdot 0) & \cdots & \cos(\Omega_L \cdot 0) & \sin(\Omega_1 \cdot 0) & \cdots & \sin(\Omega_L \cdot 0) \\ \cos(\Omega_1 \cdot 1) & \cos(\Omega_2 \cdot 1) & \cdots & \cos(\Omega_L \cdot 1) & \sin(\Omega_1 \cdot 1) & \cdots & \sin(\Omega_L \cdot 1) \\ \vdots & \vdots & & \vdots & & & \vdots \\ \cos(\Omega_1 (N-1)) & \cos(\Omega_2 (N-1)) & \cdots & \cos(\Omega_L (N-1)) & \sin(\Omega_1 (N-1)) & \cdots & \sin(\Omega_L (N-1)) \end{bmatrix}$$

$\underline{X}$ is $2L \times N$ matrix

so $A = \begin{bmatrix} 1 & \\ 1 & \\ \vdots & \underline{X} \\ 1 & \end{bmatrix}$ A is $(2L+1) \times N$ matrix

(b) No if Ω_ℓ is unknown this would not be a linear regression problem

because the relationship between y and $\cos \Omega_\ell \cdot \sin \Omega_\ell$ is not linear.

4.

(a) $\underline{\beta} = [1 \; a_1 \; a_2 \cdots a_M \; b_0 \; b_1 \; b_2 \cdots b_N]$

$M+N+2$ unknown parameters

(b) $\underline{y} \approx \underline{A}\underline{\beta}$

$\underline{y} = [y_0 \; y_1 \; y_2 \cdots y_{T-1}]$

$T \times (1+M+N+1)$

$$\underline{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & u_0 & 0 & 0 & \cdots & 0 \\ 1 & y_0 & 0 & 0 & \cdots & u_1 & u_0 & 0 & \cdots & 0 \\ 1 & y_1 & y_0 & 0 & \cdots & u_2 & u_1 & u_0 & \cdots & 0 \\ 1 & y_2 & y_1 & y_0 & \cdots & u_3 & u_2 & u_1 & u_0 & 0 \\ \vdots & & & & & & & & & \vdots \\ 1 & y_{T-2} & y_{T-3} & y_{T-4} & \cdots y_{T-1-M} & u_{T-1} & u_{T-2} & \cdots & & u_{T-1-N} \end{bmatrix}$$