In [1]:
```
#import 套件
import numpy as np
import pandas as pd
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import log_loss
```

In [2]:
```
#EXCEL做表格
data=pd.read_csv('HW3_data.csv')
```

In [3]:
```
data
```

Out[3]:

	Model1	Unnamed: 1	Probablity	Unnamed: 3	Unnamed: 4	Answer	Unnamed: 6	Unnamed: 7
0	Target	NaN	Dog	Cat	Other	Dog	Cat	Other
1	Data1	Dog	0.4	0.3	0.3	1	0	0
2	Data2	Cat	0.3	0.4	0.3	0	1	0
3	Data3	Dog	0.5	0.2	0.3	1	0	0
4	Data4	Other	0.6	0.2	0.2	0	0	1
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	Model2	NaN	Probablity	NaN	NaN	Answer	NaN	NaN
7	Target	NaN	Dog	Cat	Other	Dog	Cat	Other
8	Data1	Dog	0.8	0.1	0.1	1	0	0
9	Data2	Cat	0.1	0.7	0.2	0	1	0
10	Data3	Dog	0.7	0.1	0.2	1	0	0
11	Data4	Other	0.4	0.3	0.3	0	0	1

In [4]:
```
#雖然有缺失值和Unname不過並不影響我們計算MSE MAE 跟交換熵
```

In [5]:
```
P_model1=[0.4,0.3,0.3,0.3,0.4,0.3,0.5,0.2,0.3,0.6,0.2,0.2]
P_model2=[0.8,0.1,0.1,0.1,0.7,0.2,0.7,0.1,0.2,0.4,0.3,0.3]
```

In [6]:
```
true_model1=[1,0,0,0,1,0,1,0,0,0,0,1]
true_model2=[1,0,0,0,1,0,1,0,0,0,0,1]
```

這邊可以討論，網路說有人說僅需將分類的最高機率代入即可，但我這邊是按照老師的方法將全部代入。

In [7]:
```
mse1=mean_squared_error(true_model1,P_model1)
mse2=mean_squared_error(true_model2,P_model2)
mae1=mean_absolute_error(true_model1,P_model1)
mae2=mean_absolute_error(true_model2,P_model2)
```

In [8]:
```
print("模型1的mse:",mse1,
      "mae:",mae1)
print("模型2的mse:",mse2,
      "mae:",mae2)
```

```
模型1的mse: 0.20833333333333334 mae: 0.4166666666666666
模型2的mse: 0.09000000000000001 mae: 0.25
```

In [9]:
```
Q1=log_loss(true_model1,P_model1)
Q2=log_loss(true_model2,P_model2)
```

In [10]:
```
#跟老師的結果有出入
print("模型1的交換墒:",Q1)
print("模型2的將換墒:",Q2)
```

```
模型1的交換墒: 0.6067599259115494
模型2的將換墒: 0.32297466470683817
```

MSE、MAE 和交換墒跟老師上課所計算的不一樣，
不確定是否為 sklearn 的問題嗎?
因為我用 excel 算起來跟老師投影片上的素質是一樣的。

ANSWER			
	MSE	MAE	CROSS er
MODEL1	0.625	1.25	5.965784
MODEL 2	0.27	0.75	3.08804