In [26]:
```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [27]:
```
# 把balance欄位(y)與var1-4欄位區分開(X)
y = df.balance
X = df.drop('balance', axis=1)

# Train model
clf_0 = LogisticRegression().fit(X, y)

# Predict on training set
pred_y_0 = clf_0.predict(X)
```

X

| | var1 | var2 | var3 | var4 |
|---|---|---|---|---|
| **0** | 1 | 1 | 1 | 1 |
| **1** | 1 | 1 | 1 | 2 |
| **2** | 1 | 1 | 1 | 3 |
| **3** | 1 | 1 | 1 | 4 |
| **4** | 1 | 1 | 1 | 5 |
| **5** | 1 | 1 | 2 | 1 |
| **6** | 1 | 1 | 2 | 2 |

y

| | |
|---|---|
| 0 | 1 |
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 0 |
| 5 | 0 |
| 6 | 0 |
| 7 | 0 |
| 8 | 0 |
| 9 | 0 |
| 10 | 0 |
| 11 | 0 |

**Clf_0**：用邏輯思回歸模型分別帶入X,y運算

**Pred_y_0**：用預測函數predic，預測X每一欄 的結果是0或1

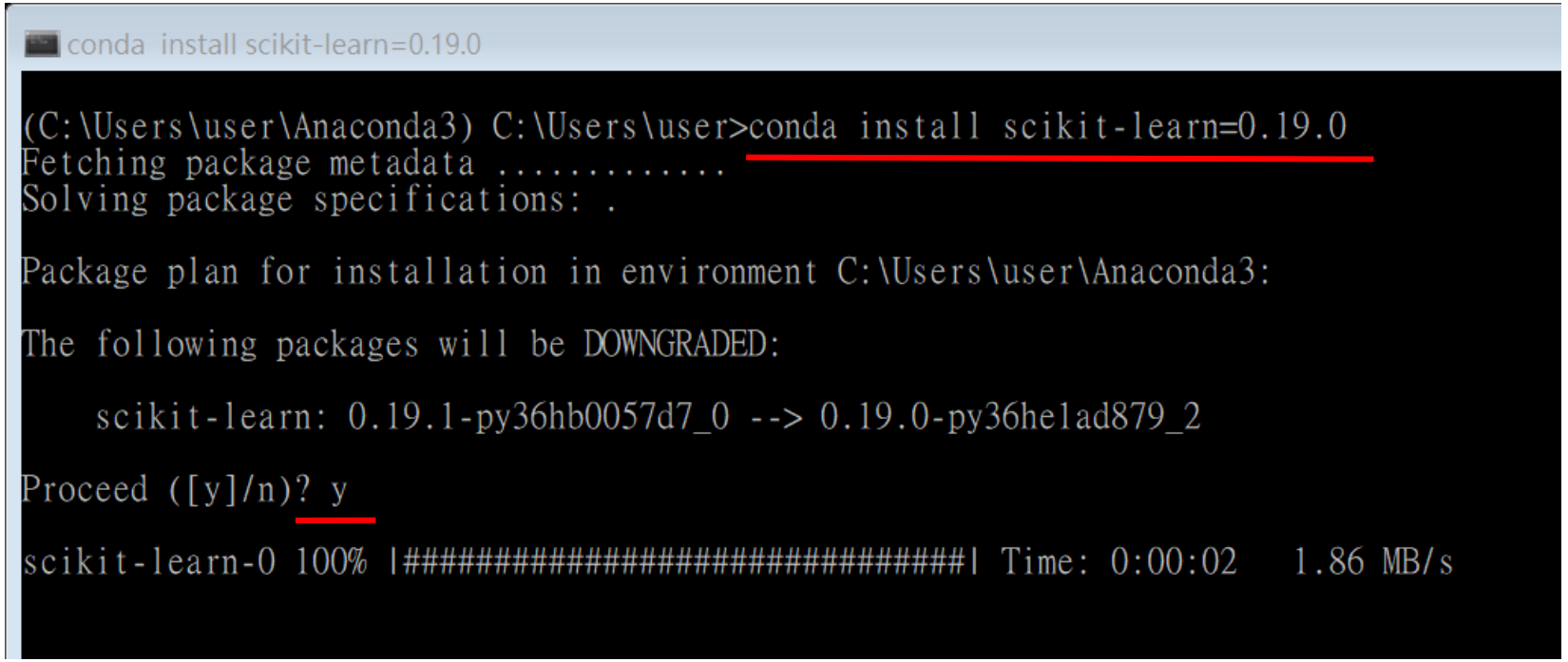
羅吉斯迴歸與線性迴歸的差別在於：前者用於**類別型**資料、後者用於連續型資料

```
import numpy as np
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import recall_score
from imblearn.over_sampling import SMOTE
```

←所需套件，
先更新↓

```
ImportError: A sklearn version of at least 0.19.0 is requir
ed to use imbalanced-learn. 0.18.1 was found. Please upgrade
sklearn
```

**Scikit-Learn-機器學習庫**
**非常實用的機器學習演算法庫，包含了基本你覺得你能用上所有機器學習演算法**



python - How to upgrade scik | IABD/ | Oversample in Predictive Mod

localhost:8888/notebooks/IABD/Oversample in Predictive Modeling.ipynb 120%

jupyter Oversample in Predictive Modeling Last Checkpoint: 22 minutes ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Code

In [1]:
```
import numpy as np
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import recall_score
from imblearn.over_sampling import SMOTE
```

In [2]:
```
import sklearn
print(sklearn.__version__)
```

```
0.19.0
```

In [ ]: