In [1]:
```
import pandas as pd
import numpy as np
from sklearn import set_config
set_config(display='diagram')
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
census = pd.read_csv('adult-census.csv')
```

In [2]:
```
target_col = 'class'
feature_col = census.drop(columns=target_col).select_dtypes(np.number).columns.values
feature_col
```

Out[2]:
```
array(['age', 'education-num', 'capital-gain', 'capital-loss',
       'hours-per-week'], dtype=object)
```

In [3]:
```
target = census[target_col]
target
```

Out[3]:
```
0         <=50K
1         <=50K
2          >50K
3          >50K
4         <=50K
          ...
48837     <=50K
48838      >50K
48839     <=50K
48840     <=50K
48841      >50K
Name: class, Length: 48842, dtype: object
```

In [4]:
```
features = census[feature_col]
features
```

Out[4]:

	age	education-num	capital-gain	capital-loss	hours-per-week
0	25	7	0	0	40
1	38	9	0	0	50
2	28	12	0	0	40
3	44	10	7688	0	40
4	18	10	0	0	30
...	...	...	...	...	...
48837	27	12	0	0	38
48838	40	9	0	0	40
48839	58	9	0	0	40
48840	22	9	0	0	20
48841	52	9	15024	0	40

48842 rows × 5 columns

In [10]:
```
x_train, x_test, y_train, y_test = train_test_split(
    features,
    target,
    random_state=123,
    test_size=.25,
    stratify=target
)
```

In [12]:
```
model = LogisticRegression()
```

In [14]:
```
model.fit(x_train, y_train)
```

Out[14]:
```
LogisticRegression()
```

In [24]:
```
print('The models accuracy score is', (model.score(x_test, y_test)*100),'%')
```

```
The models accuracy score is 80.71411022848251 %
```

In [ ]: