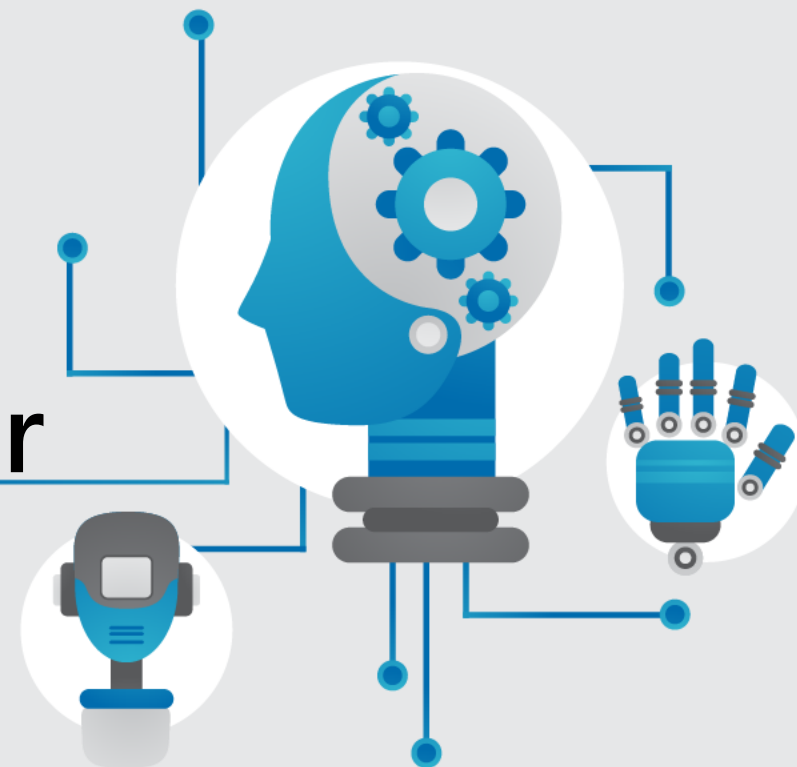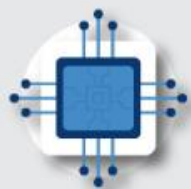# Support Vector Classifier
模型建置流程

Kaggle數據集 –
Breast Cancer Wisconsin（Diagnostic）Data Set

> 資料檔案：data.csv

> 含有569筆資料，每筆資料有32格欄位，
> 第一格為ID，第二格為labels，
> 之後的30格為該筆資料的特徵。

ID | **labels** | | | | | | **特徵**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	diagnosis	radius_mea	texture_me	perimeter_	area_mean	smoothness	compactne	concavity	concave po	symmetry	fractal_dim	radius_se	texture_se
2	842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.9053
3	842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339
4	84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869
5	84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156
6	84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813
7	843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902
8	844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732
9	84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.377
10	844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002

SVC模型建置流程

1.資料前處理
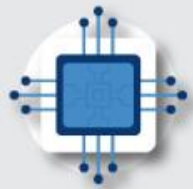
2.建構模型與參數設置

3.模型訓練與評估

4.調整模型參數

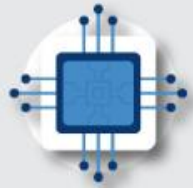**5.重複步驟2~4直到模型效率無法再改進**

資料前處理

> **從sklearn.datasets載入數據資料**

```
# 載入資料
from sklearn import datasets
bunch = datasets.load_breast_cancer()
```

資料前處理

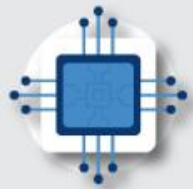› **取得data和labels，並分割成train和test資料**

```python
# 資料取得與切割
data = bunch.data
labels = bunch.target

from sklearn.model_selection import
train_test_split
X_train, X_test, Y_train, Y_test =
train_test_split(data,labels,test_size=0.3,
shuffle=True,stratify=labels)
```
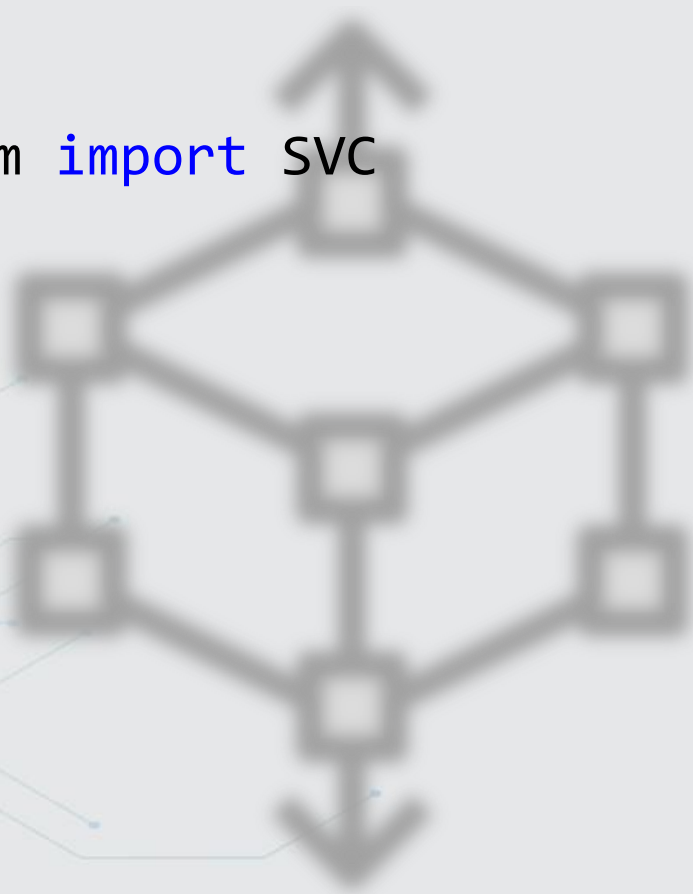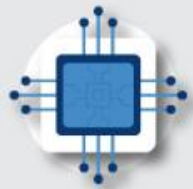
> **建構模型與參數設置**

```
# 建立模型
from sklearn.svm import SVC
svc = SVC()
```

模型訓練與評估

> **模型訓練與評估**

```python
# 進行訓練
svc.fit(X_train, Y_train)

# 進行預測
acc = svc.score(X_test, Y_test)

print('Accuracy:',acc)
```
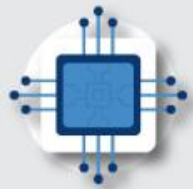
```
Accuracy: 0.6257309941520468
```

> **調整模型參數**

```
# 建立模型
from sklearn.svm import SVC
svc = SVC(C=1e5, kernel='linear',
gamma='scale', class_weight='balanced')
```

> **遞迴特徵選擇**

```python
# 用RFE,遞迴特徵選擇
from sklearn.feature_selection import RFE
selector =
RFE(estimator=svc,n_features_to_select=27)
selector = selector.fit(X_train, Y_train)
X_train = selector.transform(X_train)
X_test = selector.transform(X_test)
```

模型訓練與評估

> **模型訓練與評估**

```python
# 進行訓練
svc.fit(X_train, Y_train)

# 進行預測
acc = svc.score(X_test,
Y_test)

print('Accuracy:',acc)


Accuracy: 0.9707602339181286
```

> **將模型存於pickle檔案**

```python
import pickle

# 儲存Model
with open('svc.pickle', 'wb') as f:
    pickle.dump(svc, f)
```
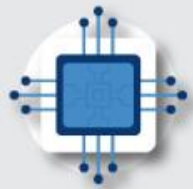
Icon made by Pixel perfect from www.flaticon.com

機器學習實務

讀取模型

> **從模型pickle檔案讀入模型**

```python
import pickle

# 讀取Model
with open('svc.pickle', 'rb') as f:
    svc = pickle.load(f)

# 測試讀入的Model
prediction = svc.predict(X_test[0:1])
```