

這段代碼主要展示了機器學習模型評估和超參數調整的最佳實踐。以下是每個主要部分的總結：

1. ****數據加載和處理****：

- 從 UCI 機器學習庫載入乳腺癌數據集。
- 使用 ``LabelEncoder`` 對標籤進行編碼，並將數據分為訓練集和測試集。

2. ****管道 (Pipeline) ****：

- 使用 ``make_pipeline`` 組合數據標準化 (``StandardScaler``)、主成分分析 (``PCA``)、和邏輯回歸 (``LogisticRegression``)。
- 評估模型的準確率。

3. ****交叉驗證****：

- 使用 ``StratifiedKFold`` 和 ``cross_val_score`` 進行 K 折交叉驗證，以評估模型的性能。
- 輸出交叉驗證準確率的均值和標準差。

4. ****學習曲線和驗證曲線****：

- 使用 ``learning_curve`` 和 ``validation_curve`` 來診斷模型的偏差和方差問題。
- 這些圖表幫助識別過擬合和欠擬合問題。

5. ****超參數調整****：

- 使用 ``GridSearchCV`` 進行網格搜尋來調整模型超參數。
- 使用 ``RandomizedSearchCV`` 和 ``HalvingRandomSearchCV`` 探索超參

數配置，以提高搜尋效率。

6. ****模型評估指標****：

- 計算和展示混淆矩陣、精確度（`precision`）、召回率（`recall`）、F1 分數（`F1 score`）和馬修斯相關係數（`MCC`）。
- 繪製接收操作特徵曲線（ROC 曲線）來評估分類模型的性能。

7. ****處理類別不平衡****：

- 顯示了如何處理類別不平衡問題，通過上採樣來平衡類別。

這些步驟和技術有助於提高模型的準確性和穩定性，並幫助選擇最適合特定數據和問題的模型。