**Logistic kfold**

# -----------------------------

# 1️⃣ Import thư viện

# -----------------------------

import numpy as np

import pandas as pd

from sklearn.datasets import load\_diabetes

from sklearn.model\_selection import StratifiedKFold, cross\_val\_score

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import make\_scorer, accuracy\_score

# -----------------------------

# 2️⃣ Load dữ liệu và chuyển target sang nhị phân

# -----------------------------

diabetes = load\_diabetes()

X = pd.DataFrame(diabetes.data, columns=diabetes.feature\_names)

# Nhị phân: 1 nếu target > median, 0 nếu <= median

y = pd.Series((diabetes.target > np.median(diabetes.target)).astype(int), name='target')

# -----------------------------

# 3️⃣ Chuẩn hóa dữ liệu

# -----------------------------

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

# -----------------------------

# 4️⃣ Khởi tạo mô hình Logistic Regression

# -----------------------------

model = LogisticRegression(max\_iter=1000)

# -----------------------------

# 5️⃣ Thiết lập k-Fold Cross Validation

# -----------------------------

k = 5

kf = StratifiedKFold(n\_splits=k, shuffle=True, random\_state=1)

# -----------------------------

# 6️⃣ Thực hiện Cross Validation

# -----------------------------

# Dùng accuracy làm metric

scores = cross\_val\_score(model, X\_scaled, y, cv=kf, scoring='accuracy')

# -----------------------------

# 7️⃣ Hiển thị kết quả

# -----------------------------

print(f"Accuracy từng fold: {scores}")

print(f"Accuracy trung bình: {scores.mean():.3f}")

print(f"Độ lệch chuẩn: {scores.std():.3f}")

**K fold linear**

# -----------------------------

# 1️⃣ Import thư viện

# -----------------------------

import numpy as np

import pandas as pd

from sklearn.datasets import fetch\_california\_housing

from sklearn.model\_selection import KFold, cross\_val\_score

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import make\_scorer, mean\_squared\_error

# -----------------------------

# 2️⃣ Load dữ liệu giá nhà

# -----------------------------

housing = fetch\_california\_housing()

X = pd.DataFrame(housing.data, columns=housing.feature\_names)

y = pd.Series(housing.target, name='MedHouseVal')

# -----------------------------

# 3️⃣ Chuẩn hóa dữ liệu

# -----------------------------

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

# -----------------------------

# 4️⃣ Khởi tạo mô hình Linear Regression

# -----------------------------

model = LinearRegression()

# -----------------------------

# 5️⃣ Thiết lập k-Fold Cross Validation

# -----------------------------

k = 5

kf = KFold(n\_splits=k, shuffle=True, random\_state=1)

# -----------------------------

# 6️⃣ Định nghĩa scorer cho MSE

# -----------------------------

mse\_scorer = make\_scorer(mean\_squared\_error, greater\_is\_better=False)

# -----------------------------

# 7️⃣ Thực hiện Cross Validation

# -----------------------------

scores = cross\_val\_score(model, X\_scaled, y, cv=kf, scoring=mse\_scorer)

# -----------------------------

# 8️⃣ Hiển thị kết quả

# -----------------------------

print(f"MSE từng fold: {-scores}") # âm vì cross\_val\_score trả âm

print(f"MSE trung bình: {-scores.mean():.3f}")

print(f"Độ lệch chuẩn: {scores.std():.3f}")