**Chương 10: Đánh Giá Mô Hình Học Máy (c10.ipynb)**

Trong chương này, tôiđã học cách đánh giá hiệu suất của các mô hình học máy bằng kỹ thuật cross-validation và các chỉ số khác nhau.

**Những Gì Đã Học Được:**

* **Cross-validation với** KFold:
  + Sử dụng KFold để chia dữ liệu thành 10 phần, với shuffle=True và random\_state=7 để đảm bảo tính ngẫu nhiên và tái lập.
* **Đánh giá mô hình phân loại**:
  + Áp dụng cross\_val\_score để tính các chỉ số như accuracy, log loss, ROC AUC.
  + Sử dụng confusion\_matrix và classification\_report để phân tích chi tiết.
* **Đánh giá mô hình hồi quy**:
  + Làm quen với MAE, MSE, và R² để đo lường hiệu suất.
* **Xử lý vấn đề hội tụ**:
  + Tăng max\_iter trong LogisticRegression để tránh cảnh báo ConvergenceWarning.

**Chương 11: Giới Thiệu Các Thuật Toán Phân Loại (c11.ipynb)**

Chương này tập trung vào việc triển khai và so sánh các thuật toán phân loại.

**Những Gì Đã Học Được:**

* **Các thuật toán phân loại**:
  + Triển khai Logistic Regression, LDA, KNN, Gaussian Naive Bayes, Decision Tree (CART), và SVM trên tập dữ liệu Pima Indians Diabetes.
* **So sánh hiệu suất**:
  + Đánh giá từng thuật toán qua cross-validation để đo độ chính xác.
* **Điều chỉnh tham số**:
  + Thử nghiệm với các tham số như max\_iter để tối ưu hóa mô hình.

**Chương 12: Giới Thiệu Các Thuật Toán Hồi Quy (c12.ipynb)**

Chương này giới thiệu các thuật toán hồi quy để dự đoán giá trị liên tục.

**Những Gì Đã Học Được:**

* **Các thuật toán hồi quy**:
  + Áp dụng Linear Regression, Ridge, Lasso, ElasticNet, KNN Regressor, Decision Tree Regressor, và SVR trên tập dữ liệu Housing.
* **Đánh giá hiệu suất**:
  + Sử dụng neg\_mean\_squared\_error để so sánh sai số giữa các mô hình.
* **Regularization**:
  + Hiểu cách Ridge, Lasso, và ElasticNet giảm overfitting.

**Chương 13: So Sánh Các Thuật Toán Phân Loại (c13.ipynb)**

Chương này giúp bạn so sánh các thuật toán phân loại một cách có hệ thống.

**Những Gì Đã Học Được:**

* **Xây dựng pipeline so sánh**:
  + Đánh giá nhiều mô hình phân loại trên cùng tập dữ liệu qua cross-validation.
* **Trực quan hóa**:
  + Vẽ boxplot bằng matplotlib để so sánh phân bố accuracy của các mô hình.
* **Phân tích kết quả**:
  + Xác định mô hình tốt nhất dựa trên hiệu suất và độ ổn định.

**Chương 14: Sử Dụng Pipeline Trong scikit-learn (c14.ipynb)**

Chương này tập trung vào việc tự động hóa quy trình học máy bằng pipeline.

**Những Gì Đã Học Được:**

* **Tạo pipeline**:
  + Kết hợp StandardScaler và mô hình như LinearDiscriminantAnalysis trong một pipeline.
* **Trích xuất đặc trưng**:
  + Sử dụng FeatureUnion để kết hợp PCA và SelectKBest.
* **Đánh giá pipeline**:
  + Dùng cross\_val\_score để đo hiệu suất của toàn bộ pipeline.

**Kết Luận**

Qua 5 chương từ 10 đến 14:

* **Chương 10**: Nắm vững kỹ thuật đánh giá mô hình.
* **Chương 11**: Làm quen với các thuật toán phân loại.
* **Chương 12**: Hiểu và áp dụng các thuật toán hồi quy.
* **Chương 13**: So sánh và trực quan hóa hiệu suất mô hình.
* **Chương 14**: Tự động hóa quy trình học máy bằng pipeline.

Những kỹ năng này giúp tôi xây dựng nền tảng vững chắc để giải quyết các bài toán học máy thực tế.