Bài tập Thực hành Python
Bài số 20: Thiết kế hướng đối tượng

Hà Minh Tuấn
Khoa Toán - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

Ngày 3 tháng 12 năm 2025

Bài tập Python OOP cho Khoa học dữ liệu

Dưới đây là 10 chủ đề bài tập thực hành thiết kế và triển khai class trong Python. Mỗi bài tập yêu cầu sinh viên xác định các thuộc tính và phương thức cần thiết, thiết kế class cha, class con, sử dụng thuộc tính bảo vệ và riêng tư, phương thức tính toán và phương thức trừu tượng.

Bài 1: Quản lý tập dữ liệu (Dataset Management)

- Yêu cầu: Thiết kế một class cha `Dataset` với các thuộc tính như tên dataset, số lượng mẫu, số lượng đặc trưng. Tạo các class con `CSVData`, `SQLData`, mỗi class con phải triển khai phương thức trừu tượng `load_data()`.
- Sinh viên phải sử dụng: thuộc tính riêng tư cho đường dẫn file, thuộc tính tính toán để trả về số lượng bản ghi, và phương thức hiển thị thông tin dataset.

Bài 2: Tiền xử lý dữ liệu (Data Preprocessing)

- Yêu cầu: Class cha `Preprocessor` với phương thức trừu tượng `process(data)`. Class con `Normalizer`, `Standardizer`, `Imputer` triển khai các phương thức cụ thể.
- Sinh viên cần sử dụng: thuộc tính bảo vệ để lưu thông tin trạng thái (fit/transform), thuộc tính tính toán để trả về số lượng giá trị bị thiếu sau khi xử lý.

Bài 3: Quản lý mô hình học máy (ML Model Management)

- Yêu cầu: Class cha `MLModel` với các thuộc tính như tên mô hình, siêu tham số, trạng thái huấn luyện. Class con `LinearRegressionModel`, `RandomForestModel` triển khai phương thức trừu tượng `train()` và `predict()`.
- Sinh viên phải sử dụng thuộc tính riêng tư cho trọng số mô hình, thuộc tính tính toán để đánh giá accuracy trên tập validation.

Bài 4: Đánh giá mô hình (Model Evaluation)

- Yêu cầu: Class cha `Evaluator` với phương thức trừu tượng `evaluate(model, data)`. Class con `ClassificationEvaluator`, `RegressionEvaluator` triển khai các chỉ số như accuracy, RMSE, R2.

- Sinh viên cần sử dụng thuộc tính bảo vệ để lưu kết quả trung gian và thuộc tính tính toán để trả về các metric.

Bài 5: Quản lý pipeline xử lý dữ liệu (Data Pipeline)

- Yêu cầu: Class `PipelineStep` (cha) với phương thức trừu tượng `execute(data)`. Các class con như `CleaningStep`, `FeatureSelectionStep`, `ModelTrainingStep` triển khai phương thức `execute`.
- Sinh viên phải sử dụng: thuộc tính riêng tư để lưu trạng thái thực thi, phương thức tính toán để báo cáo tiến trình.

Bài 6: Quản lý thí nghiệm máy học (Experiment Tracking)

- Yêu cầu: Class cha `Experiment` với thuộc tính tên, ngày tạo, mô hình sử dụng, dữ liệu sử dụng. Class con `ClassificationExperiment`, `RegressionExperiment` triển khai phương thức trừu tượng `run()`.
- Sinh viên cần dùng thuộc tính bảo vệ để lưu log thí nghiệm và phương thức tính toán để tổng hợp kết quả tốt nhất.

Bài 7: Phân loại văn bản (Text Classification)

- Yêu cầu: Class cha `TextProcessor` với phương thức trừu tượng `tokenize(text)`. Class con `TFIDFProcessor`, `Word2VecProcessor` triển khai các phương thức đặc thù.
- Sinh viên phải dùng thuộc tính riêng tư để lưu bộ từ vựng, thuộc tính tính toán để trả về số lượng token.

Bài 8: Quản lý tập hợp đặc trưng (Feature Set Management)

- Yêu cầu: Class cha `FeatureSet` với phương thức trừu tượng `compute()` để tính toán giá trị đặc trưng. Class con `NumericFeatureSet`, `CategoricalFeatureSet` triển khai phương thức `compute()` riêng.
- Sinh viên cần sử dụng thuộc tính bảo vệ để lưu kết quả tính toán tạm thời và thuộc tính tính toán để trả về số lượng đặc trưng được xử lý.

Bài 9: Mô phỏng dữ liệu sinh học (Biological Data Simulation)

- Yêu cầu: Class cha `Simulator` với phương thức trừu tượng `simulate(n)`. Class con `GeneExpressionSimulator`, `ProteinInteractionSimulator` triển khai phương thức mô phỏng dữ liệu.
- Sinh viên sử dụng thuộc tính riêng tư để lưu tham số mô phỏng và thuộc tính tính toán để thống kê dữ liệu sinh ra.

Bài 10: Quản lý kết quả phân tích (Analysis Result Management)

- Yêu cầu: Class cha `AnalysisResult` với phương thức trừu tượng `summarize()`. Class con `StatisticalResult`, `MLResult` triển khai phương thức tóm tắt kết quả.
- Sinh viên cần sử dụng thuộc tính bảo vệ để lưu dữ liệu trung gian, thuộc tính tính toán để xuất báo cáo tự động.