

Học Máy: Có Giám Sát và Không Giám Sát

I. Học Có Giám Sát (Supervised Learning)



Định nghĩa

Dữ liệu huấn luyện có nhãn.



Mục tiêu

Học mối quan hệ Input (X) → Output (Y).



Ứng dụng

Dự đoán nhãn cho dữ liệu mới.

Quy Trình Supervised Learning

1

Thu thập & Gán nhãn

Data có label.

2

Tiền xử lý

Làm sạch, chuẩn hóa.

3

Chia tập dữ liệu

Train, Validation, Test.

4

Huấn luyện mô hình

Fit trên tập huấn luyện.

5

Đánh giá & Tối ưu

Metrics, Hyperparameter Tuning.

6

Triển khai & Giám sát

Trong sản phẩm, cập nhật.

Classification vs. Regression

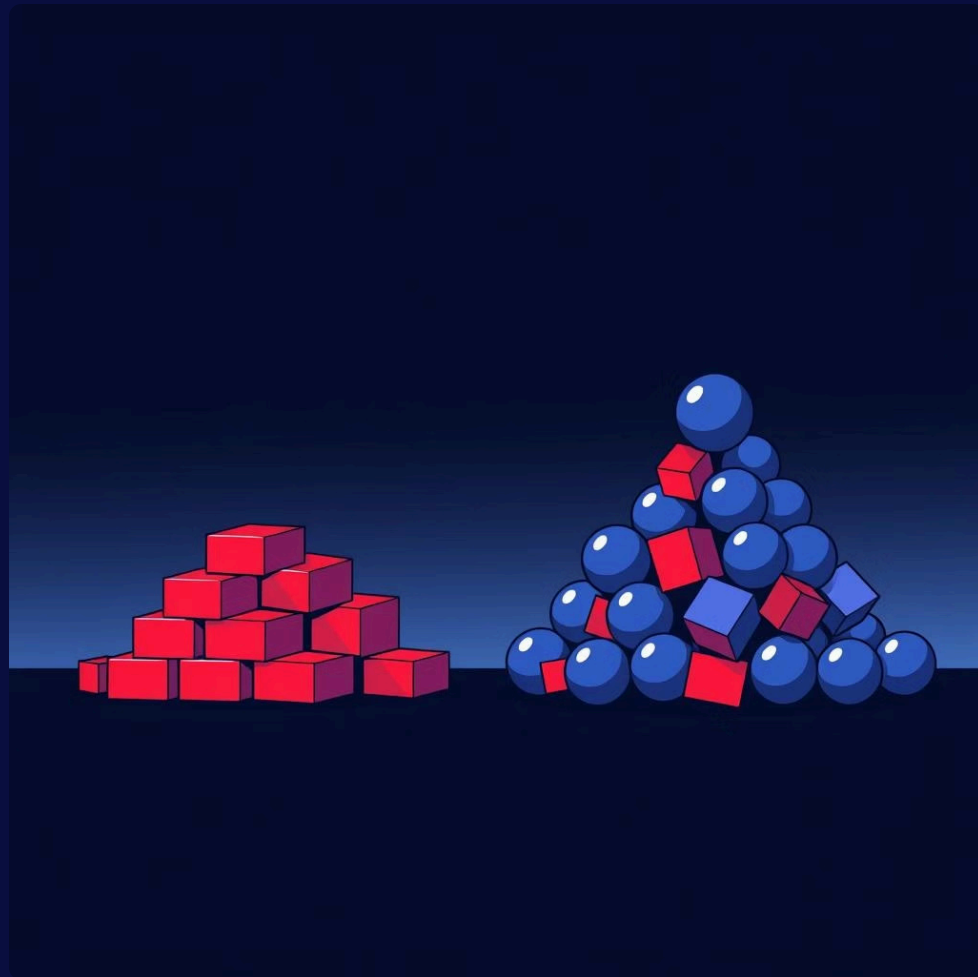
Classification

Nhãn rời rạc (0/1, A/B/C).

- Dự đoán spam
- Phân loại ảnh

Đánh giá: Accuracy, F1-score.

Thuật toán: Logistic Regression, SVM.



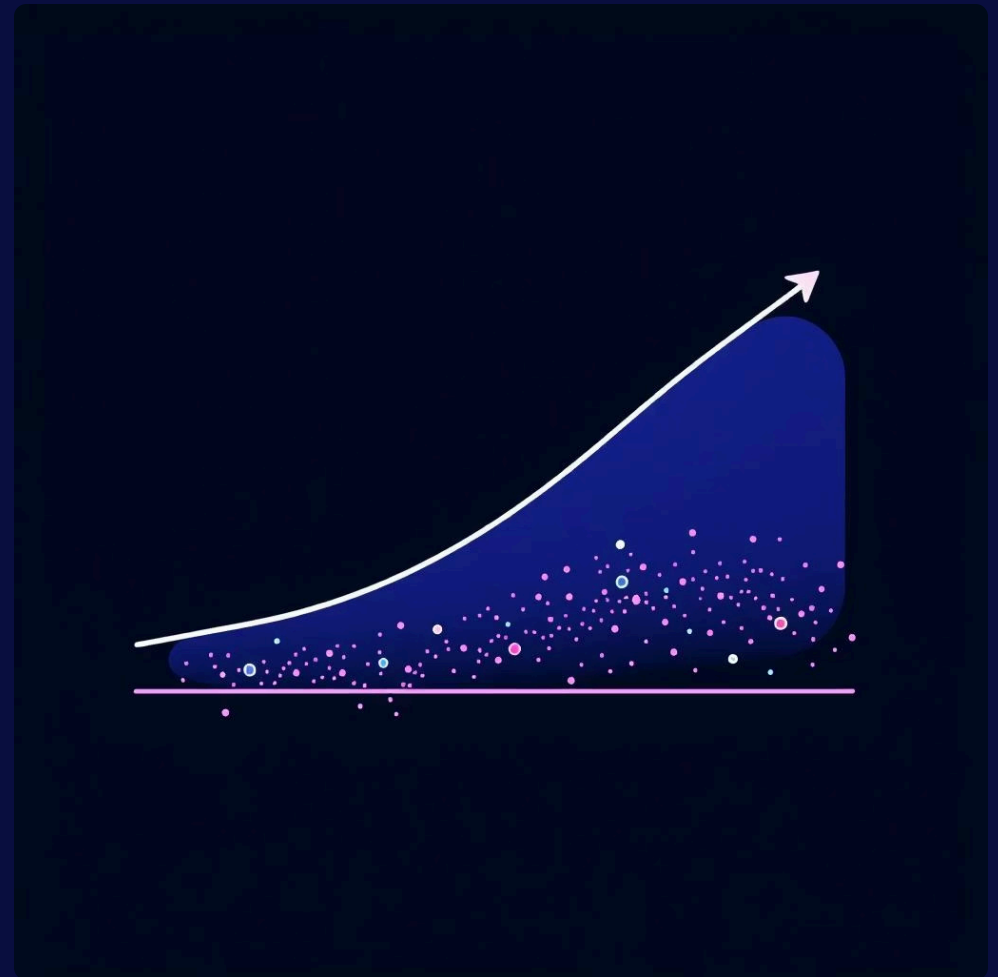
Regression

Giá trị liên tục.

- Dự đoán giá nhà
- Dự đoán điểm số

Đánh giá: MSE, MAE.

Thuật toán: Linear Regression, SVR.



II. Học Không Giám Sát (Unsupervised Learning)



Định nghĩa

Dữ liệu không có nhãn.



Mục tiêu

Tìm cấu trúc ẩn, mẫu.



Ứng dụng

Phân nhóm, giảm chiều.

Mục Tiêu & Ứng Dụng

Tìm nhóm (Cluster)

Phân nhóm khách hàng.

Tìm cấu trúc ẩn

Gợi ý sản phẩm.

Giảm chiều dữ liệu

Trực quan hóa 2D/3D.

Phát hiện bất thường

Giao dịch gian lận.



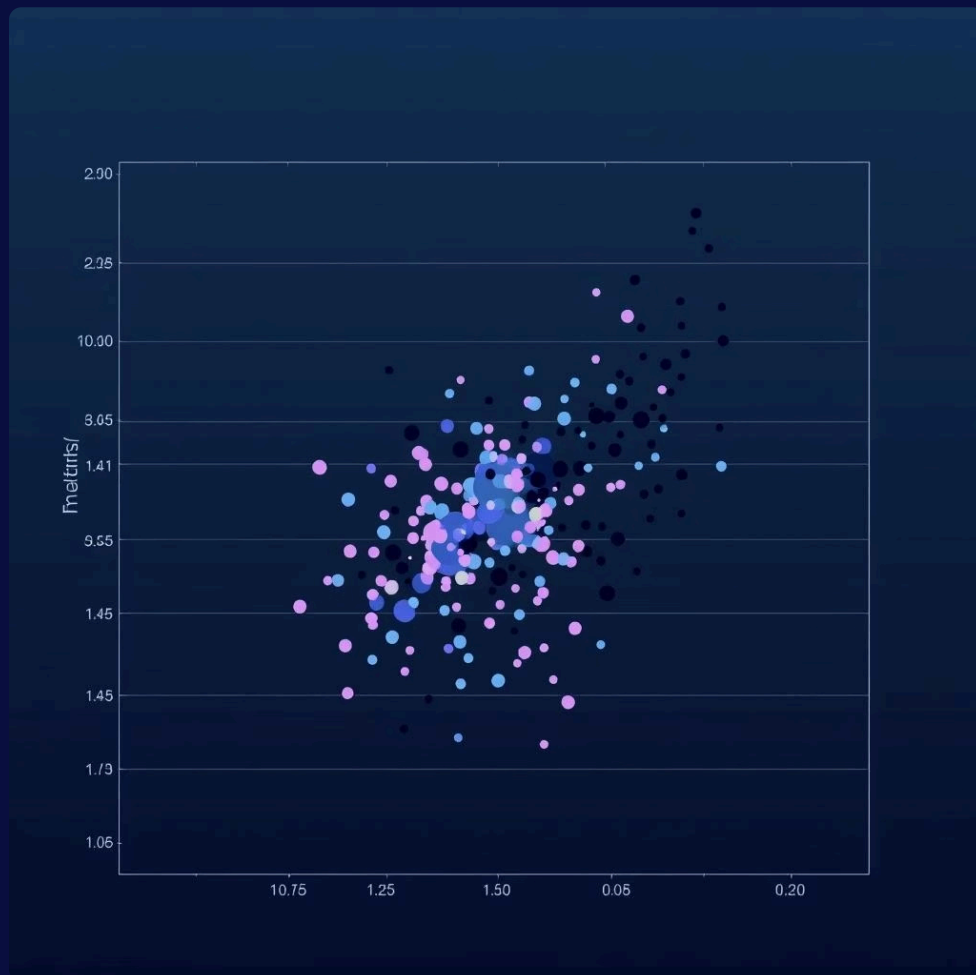
Clustering vs. Giảm Chiều

Clustering

Gom nhóm điểm giống nhau.

- Nhãn cluster
- Phân nhóm khách hàng

Thuật toán: K-Means, DBSCAN.

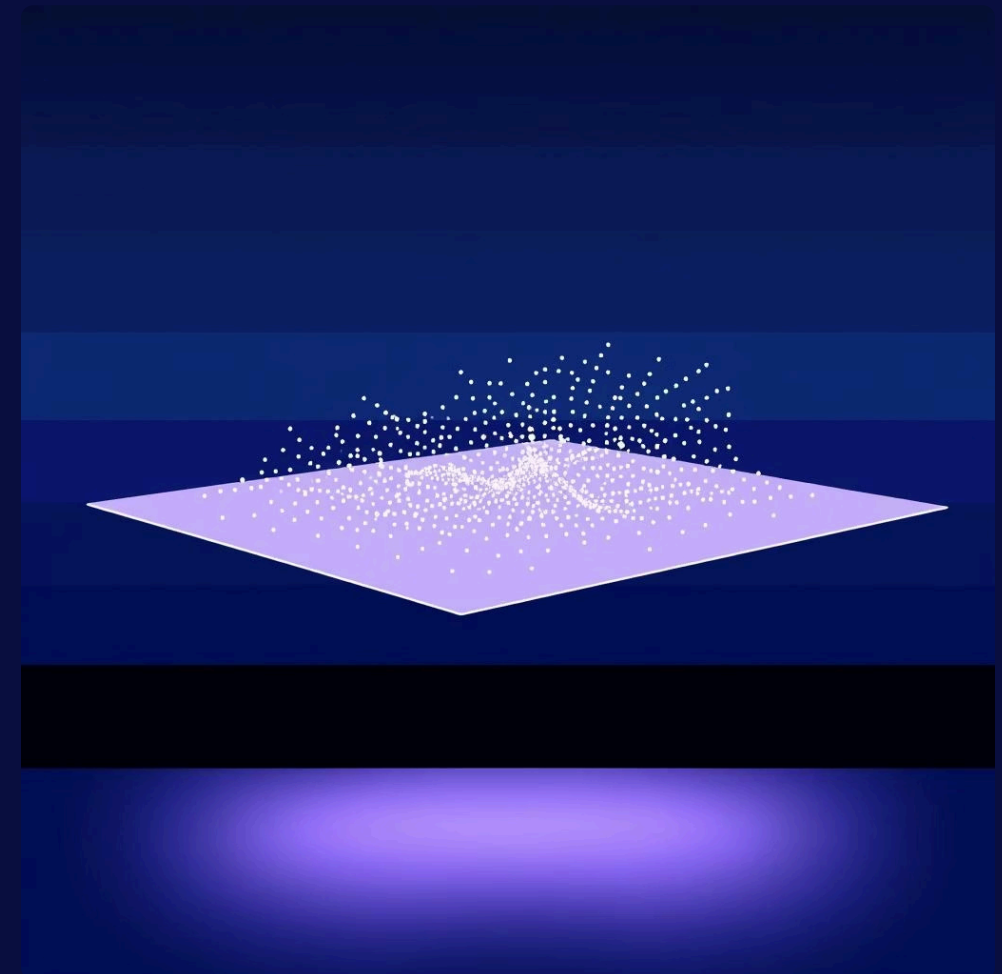


Dimensionality Reduction

Giảm số chiều, giữ thông tin.

- Dữ liệu mới (ít chiều)
- Trực quan hóa 2D

Thuật toán: PCA, t-SNE.



So Sánh Chung

Tiêu chí	Supervised Learning	Unsupervised Learning
Dữ liệu	Có nhãn	Không có nhãn
Mục tiêu	Dự đoán nhãn	Tìm cấu trúc ẩn
Khó khăn	Gán nhãn tốn kém	Khó đánh giá kết quả

Thuật Toán Supervised Learning



Linear Regression

Dự đoán giá trị liên tục.



Logistic Regression

Phân loại nhị phân.



Decision Tree

Chia dữ liệu theo cây.



Random Forest

Nhiều Decision Tree.



SVM

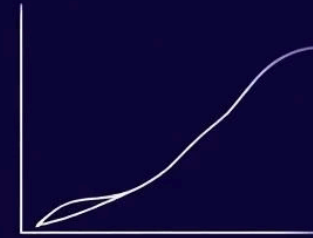
Tìm siêu phẳng phân tách.



KNN

Dựa vào hàng xóm gần nhất.

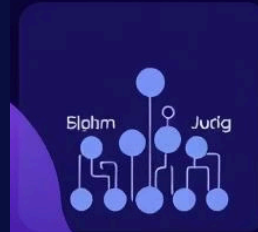
supervised learning Learning



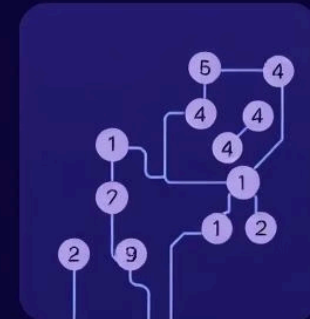
1
Linear Regression



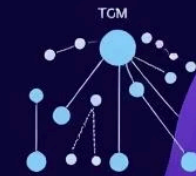
Logistic
Regression



Decision Tree
Random Forest



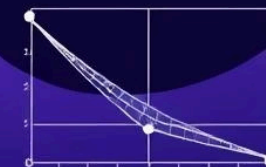
Re-forest
Random Forest



Support Vector Machine
with a hyperplane



Support
Vector Machine



SVM
Support Vector Machine



KNN
K-Nearest Neighbors

Thuật Toán Unsupervised Learning



K-Means

Gom nhóm theo khoảng cách.



DBSCAN

Gom nhóm theo mật độ.



Hierarchical Clustering

Gom nhóm phân cấp.



PCA

Giảm chiều tuyến tính.



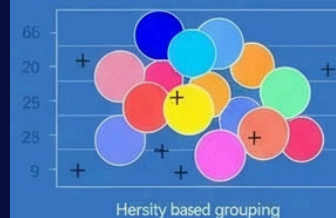
t-SNE

Giảm chiều phi tuyến.

Unsupervised Learning

Unsupervised learning is a type of machine learning where the algorithm learns from data without any supervision or labels. It is used to find hidden patterns or structures in the data.

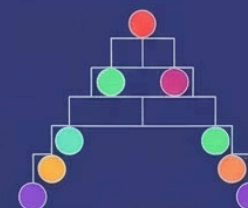
K-MEANS CLUSTERING



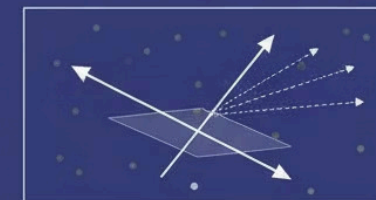
DBSCAN



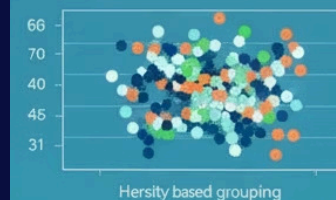
HICARIACHCAL CLUSTERING



PCA (DIMENSION REDUCTION)



t-SNE (VISULIZATION)



T-SNE

