MACHINE LEARNING

SUPERVISED LEARNING

UNSUPERVISED LEARNING

CLASSIFICATION

REGRESSION

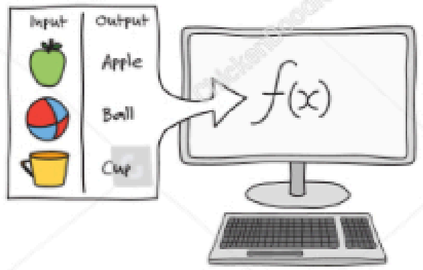
CLUSTERING

ASSOCIATION

Supervised Learning & Unsupervised Learning

Supervised Machine Learning

Example application Training a computer to recognize and classify similar objects based on shape.



The computer is given examples of inputs and typical outputs which it uses to develop and refine an algorithm.

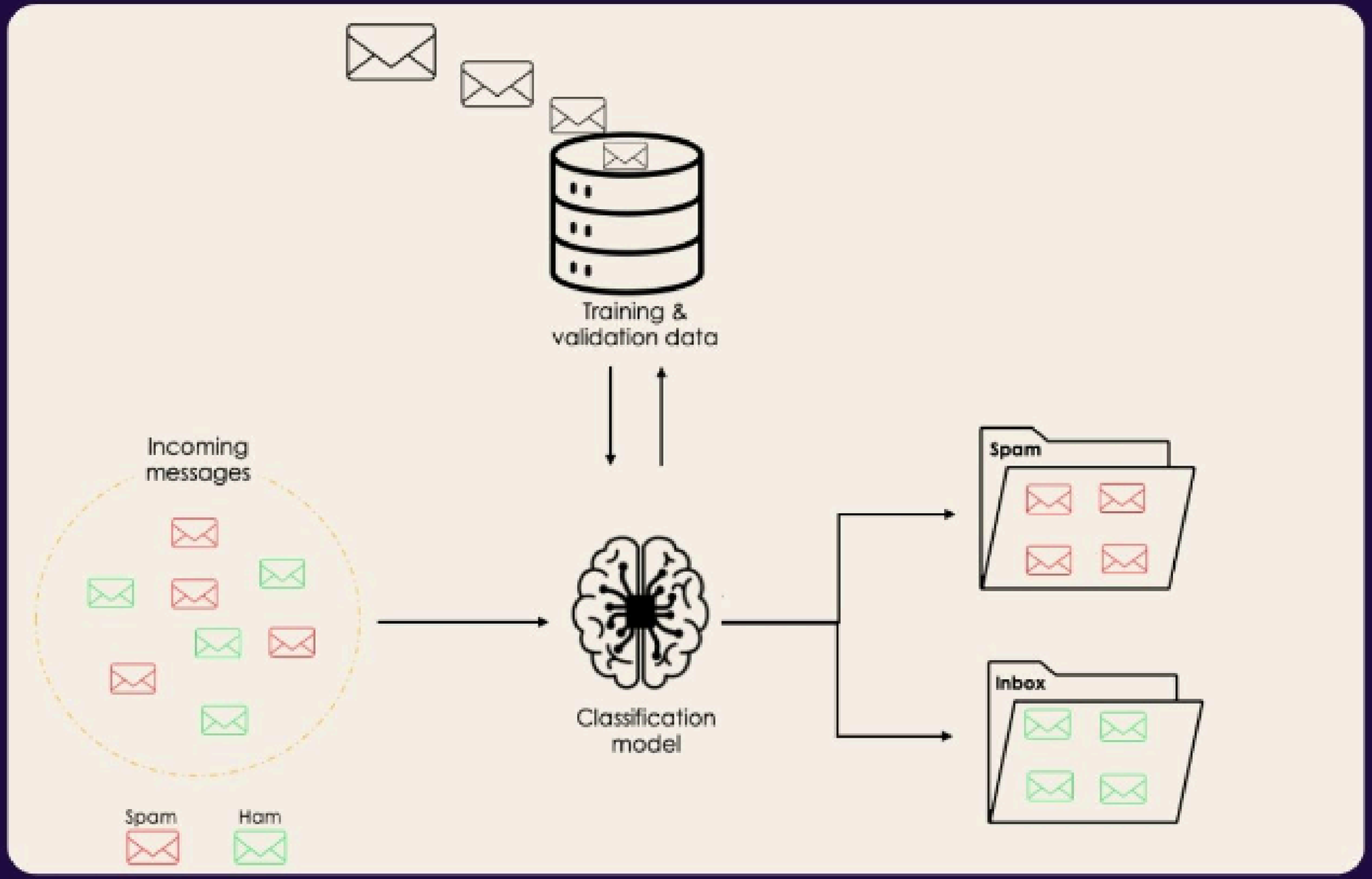
The algorithm is applied to new data and the outcome is used for further refinement.

Unsupervised Machine Learning

Example application Spotting patterns in customer data based on purchasing behaviour.

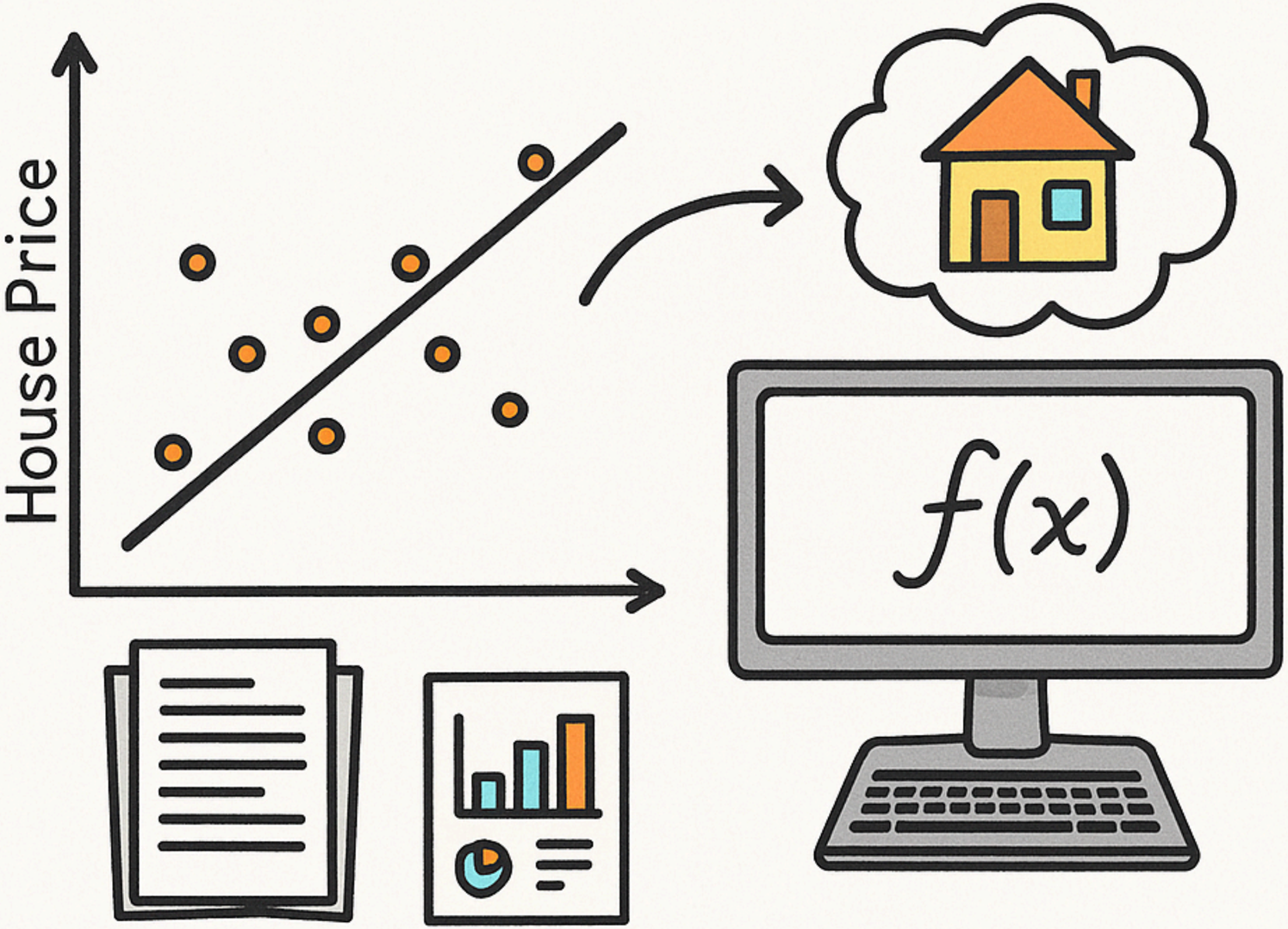
The computer learns by exploring inputted data and finding structure and data patterns on its own.

Classification in Machine Learning



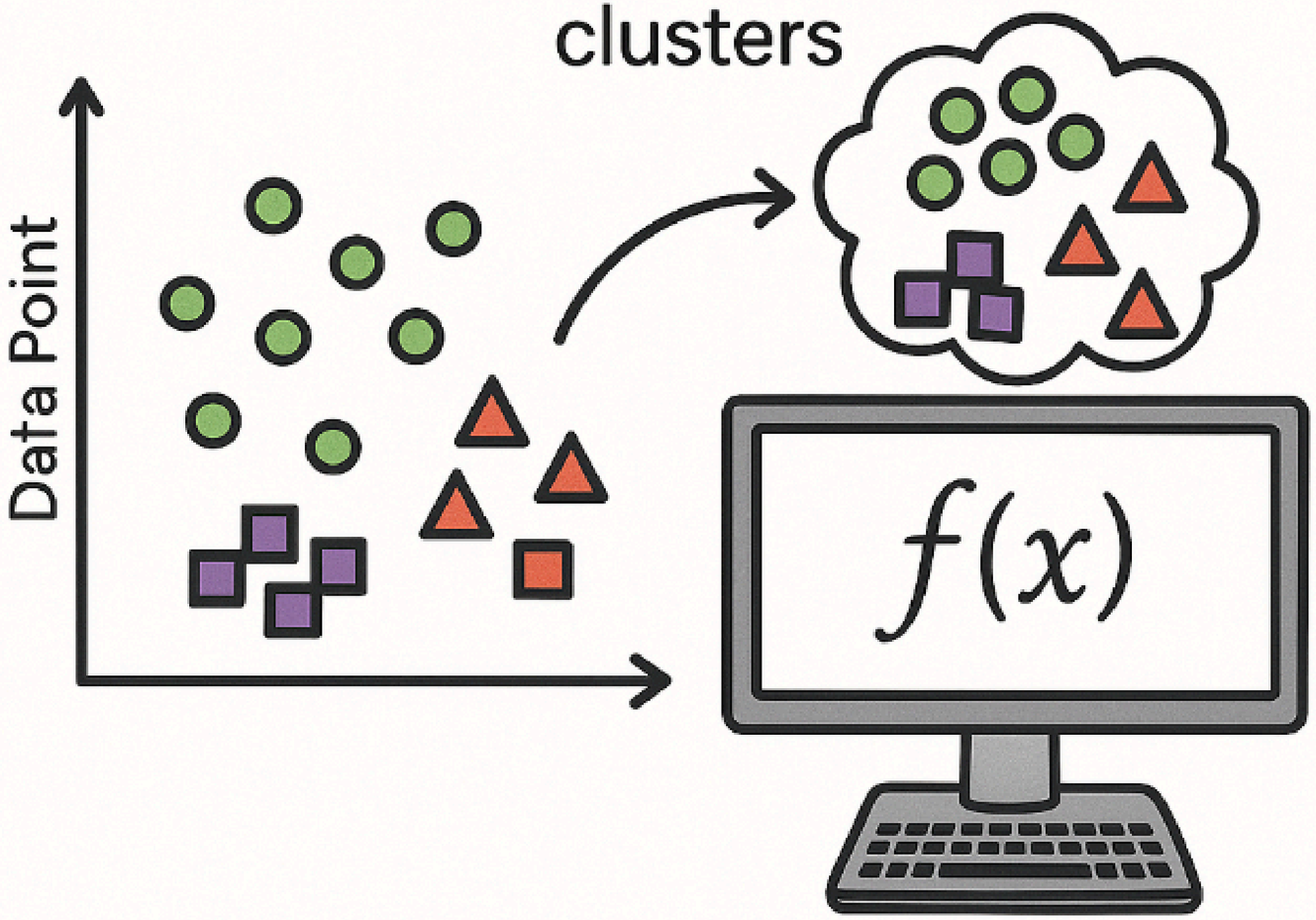
Regression

predicting house prices



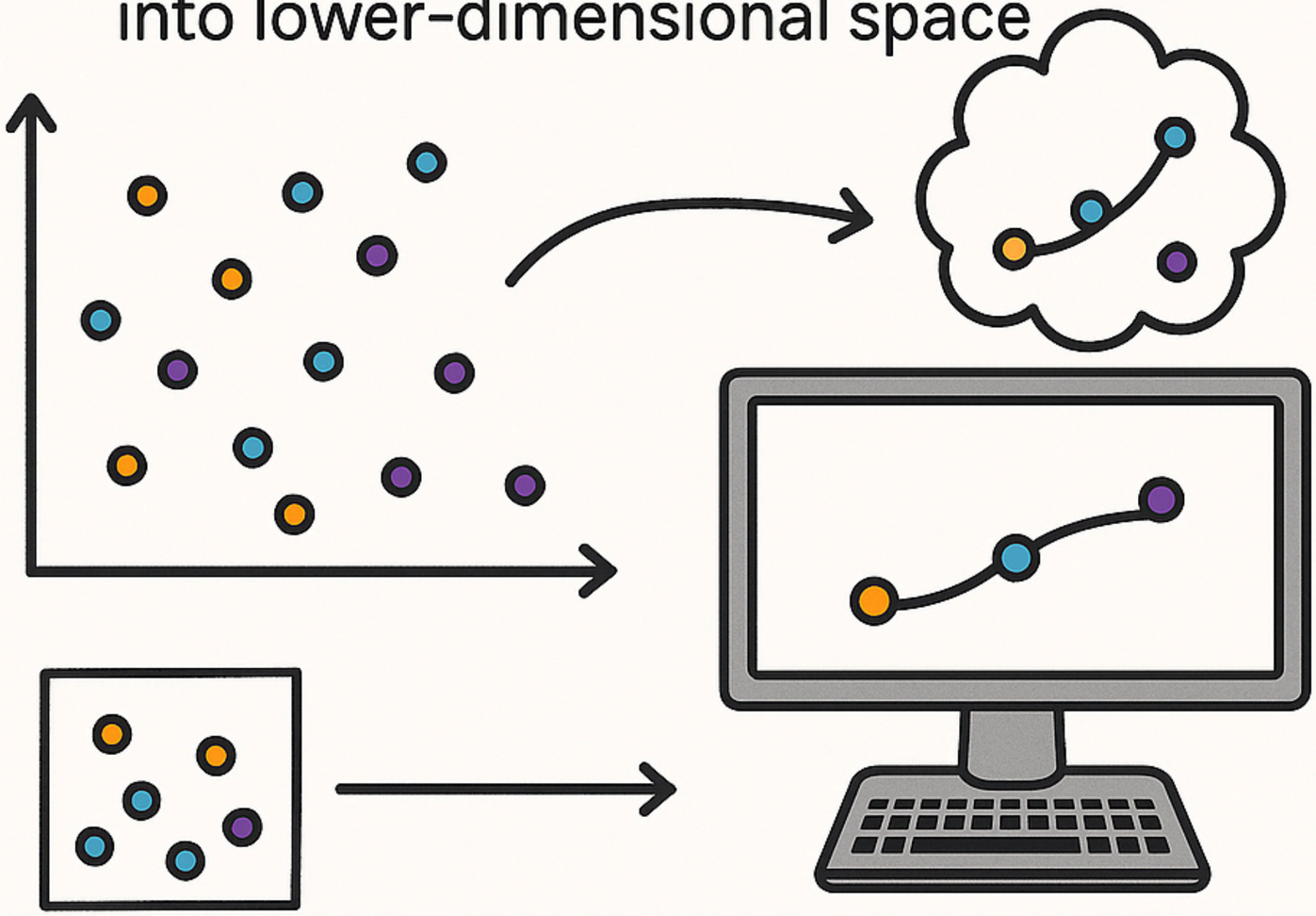
Clustering

grouping data points into clusters



Dimensionality Reduction

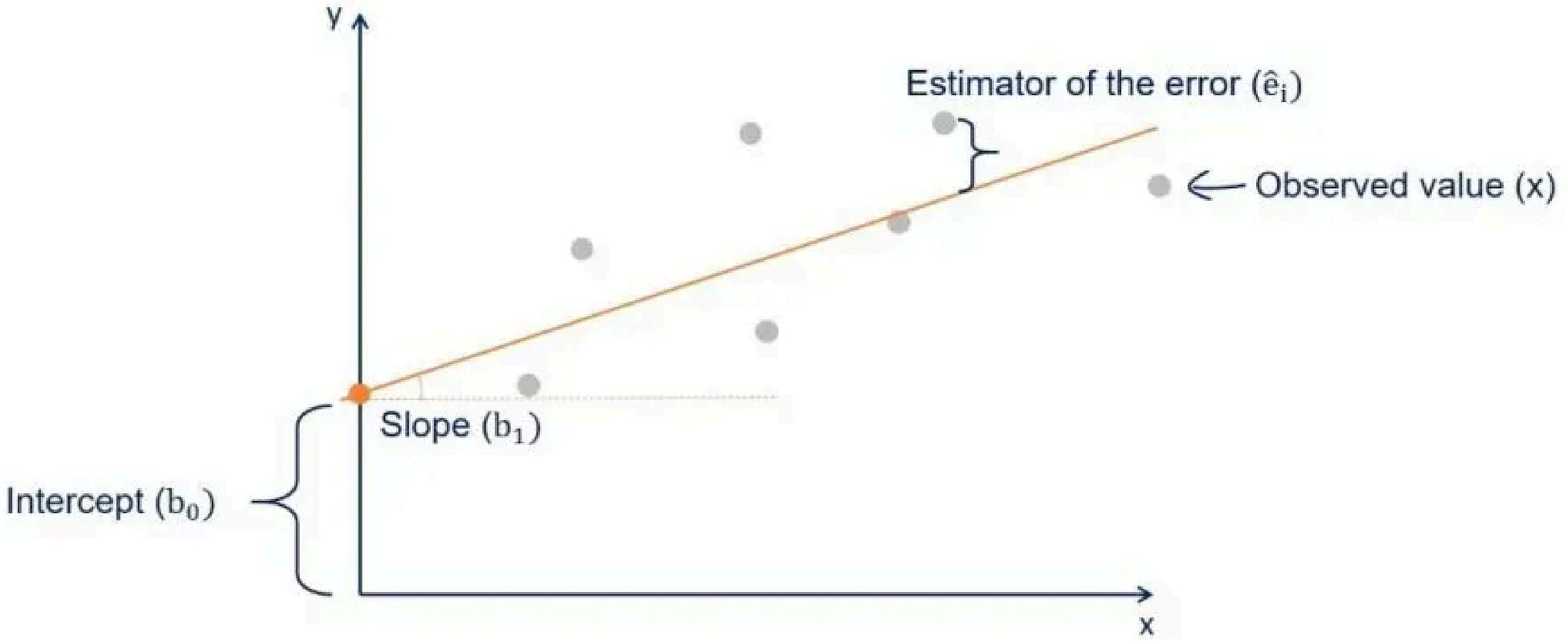
transforming high-dimensional data
into lower-dimensional space

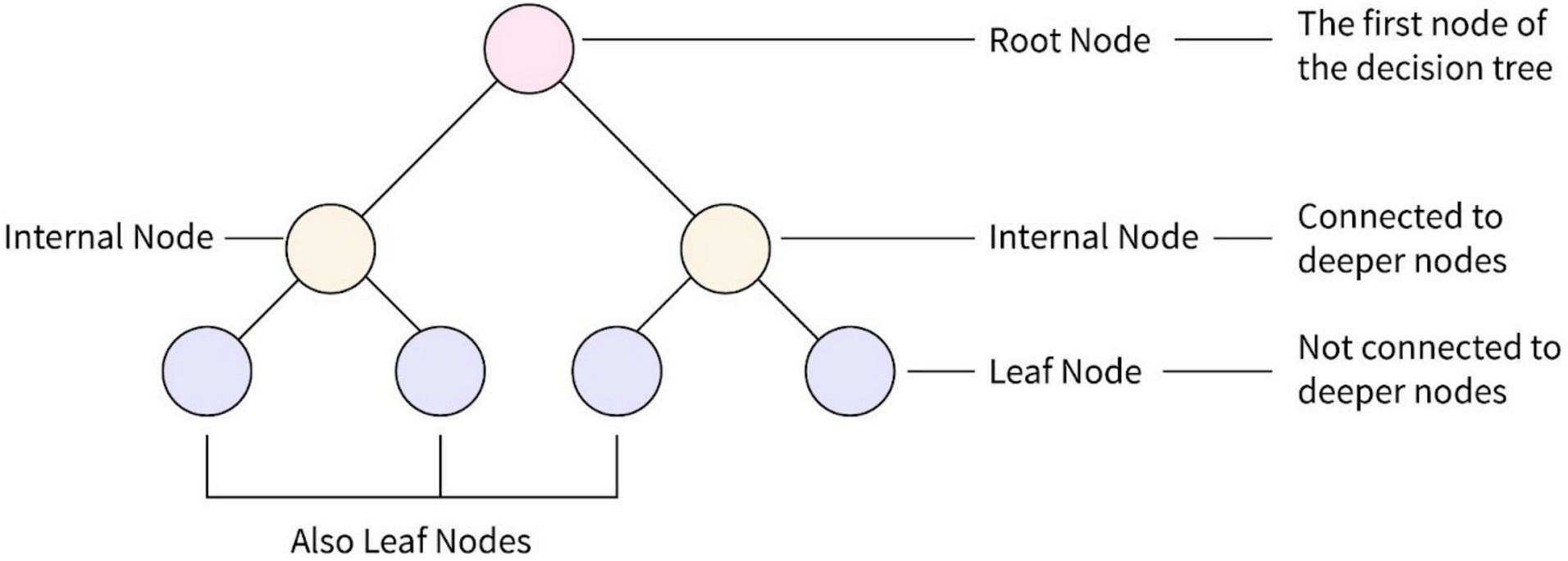


Tiêu chí	Supervised Learning (Học có giám sát)	Unsupervised Learning (Học không giám sát)
Dữ liệu	Dữ liệu đã gán nhãn	Dữ liệu chưa gán nhãn
Mục tiêu	Dự đoán hoặc phân loại đầu ra	Khám phá cấu trúc, nhóm, mẫu ẩn trong dữ liệu
Thuật toán phổ biến	SVM, Decision Tree, Linear/Logistic Regression, NN	K-means, Hierarchical Clustering, PCA, Apriori
Ứng dụng	Nhận diện hình ảnh, phát hiện spam, dự đoán giá	Phân nhóm khách hàng, phát hiện bất thường, gợi ý
Độ phức tạp	Thường đơn giản hơn, dễ đánh giá	Thường phức tạp hơn, khó đánh giá
Kết quả	Độ chính xác cao nếu dữ liệu nhãn tốt	Thường ít chính xác hơn, phụ thuộc vào thuật toán
Yêu cầu dữ liệu	Cần nhiều dữ liệu đã gán nhãn	Không cần nhãn, phù hợp dữ liệu lớn, chưa phân loại

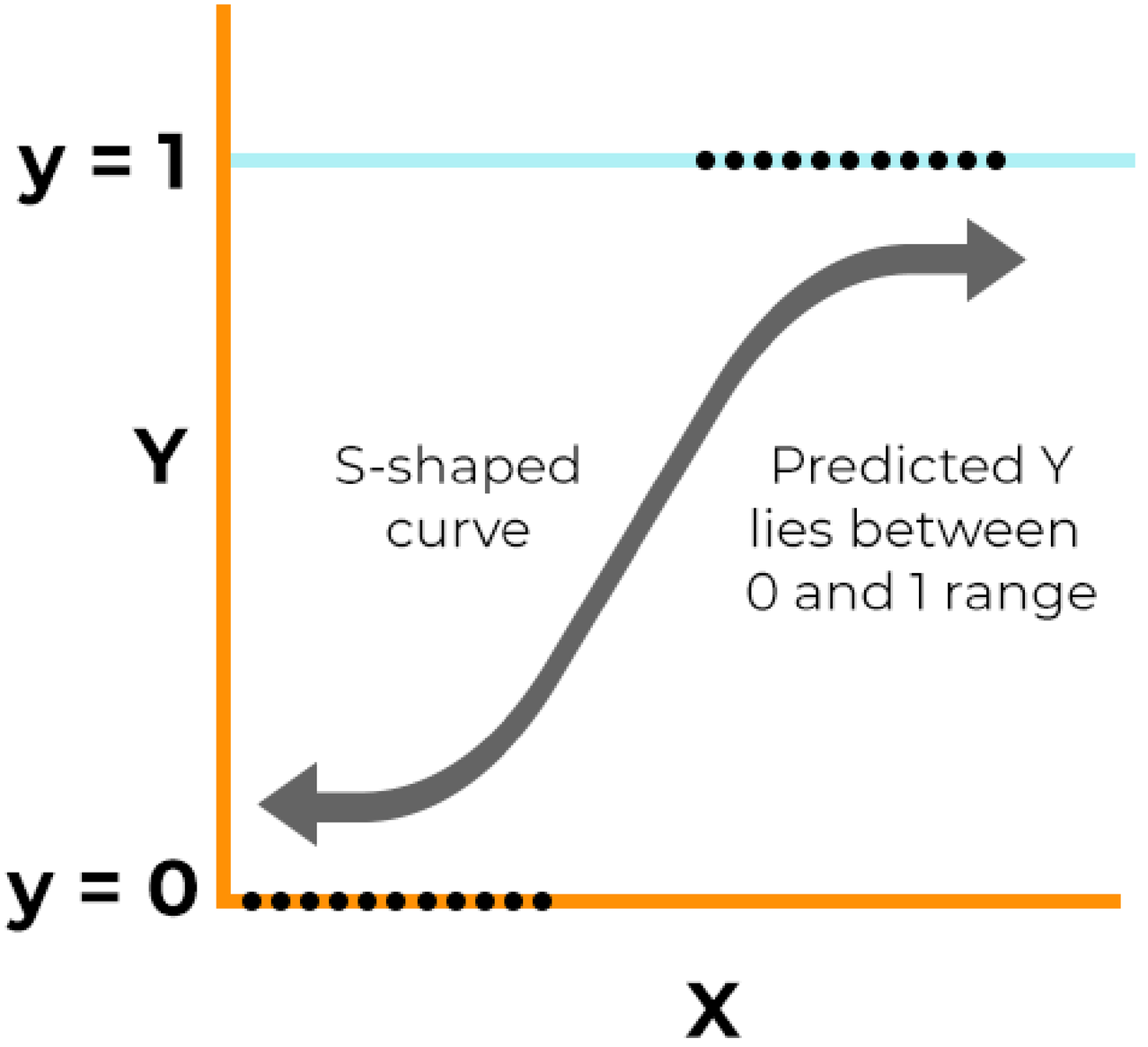
Linear regression model. Geometrical representation

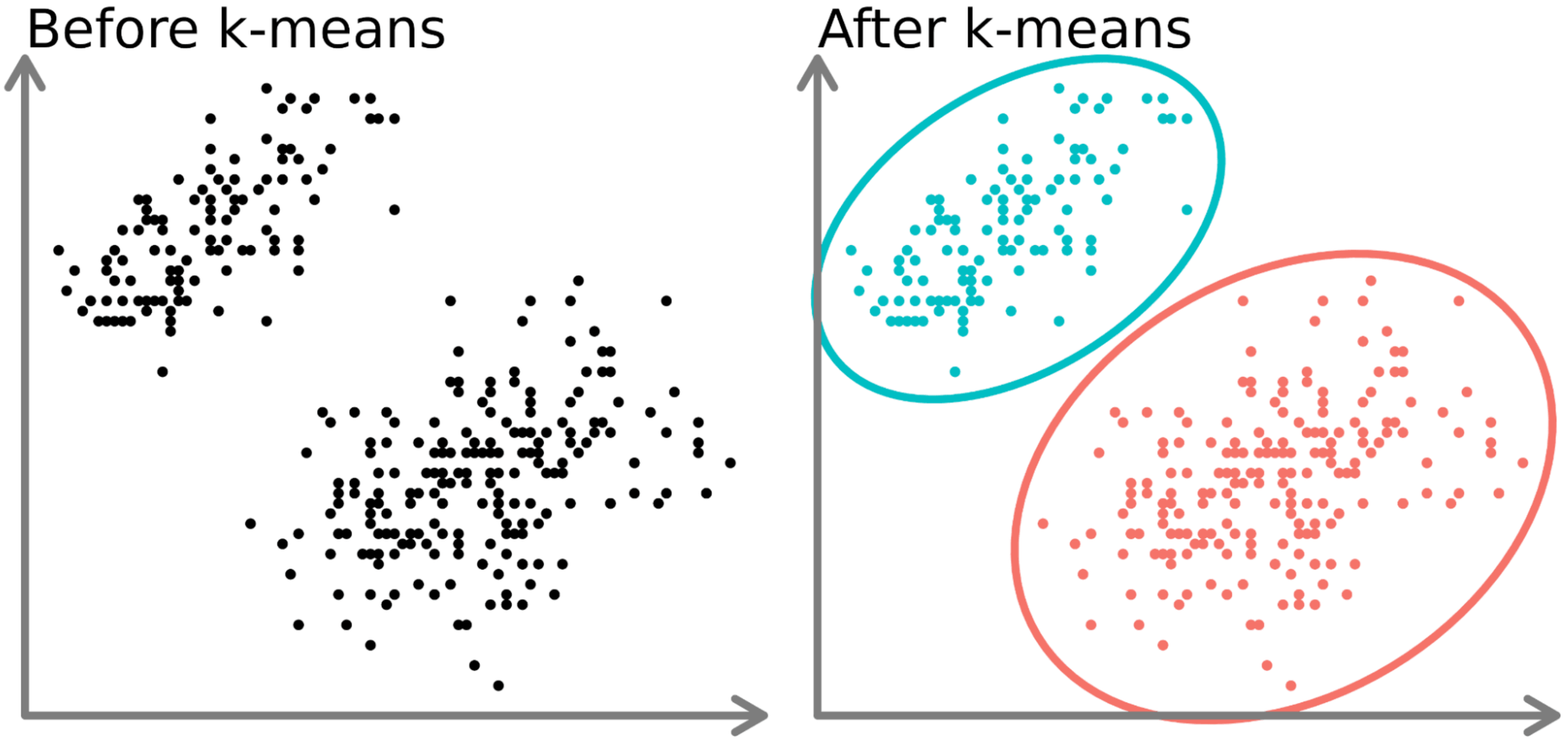
$$\hat{y}_i = b_0 + b_1 x_i$$



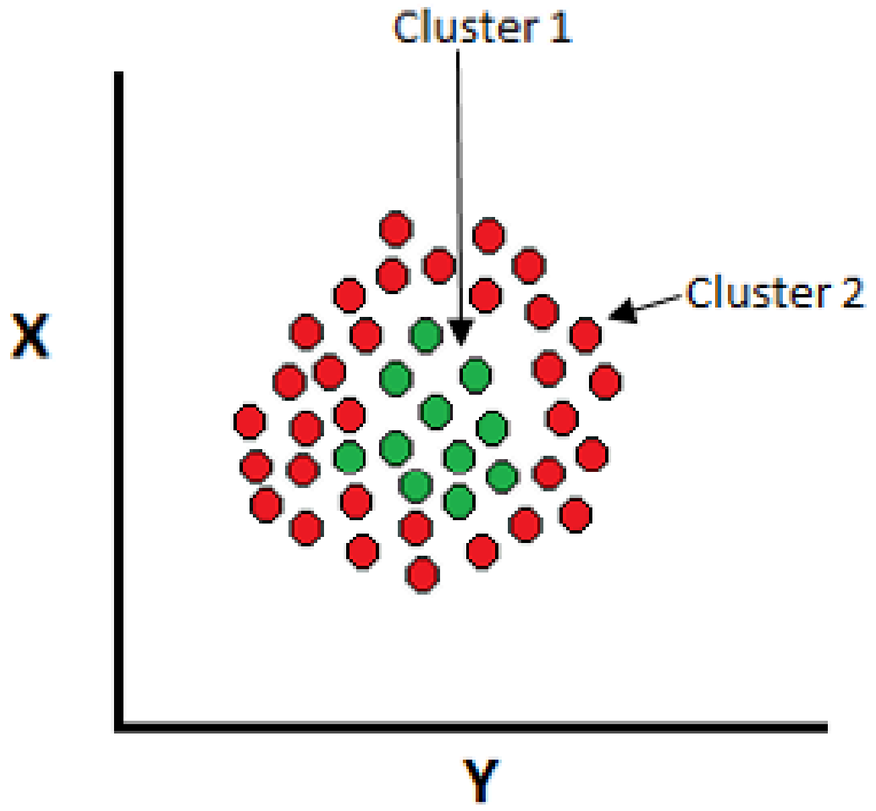

Root Node
The first node of the decision tree
Internal Node
Internal Node
Connected to deeper nodes
Leaf Node
Not connected to deeper nodes
Also Leaf Nodes

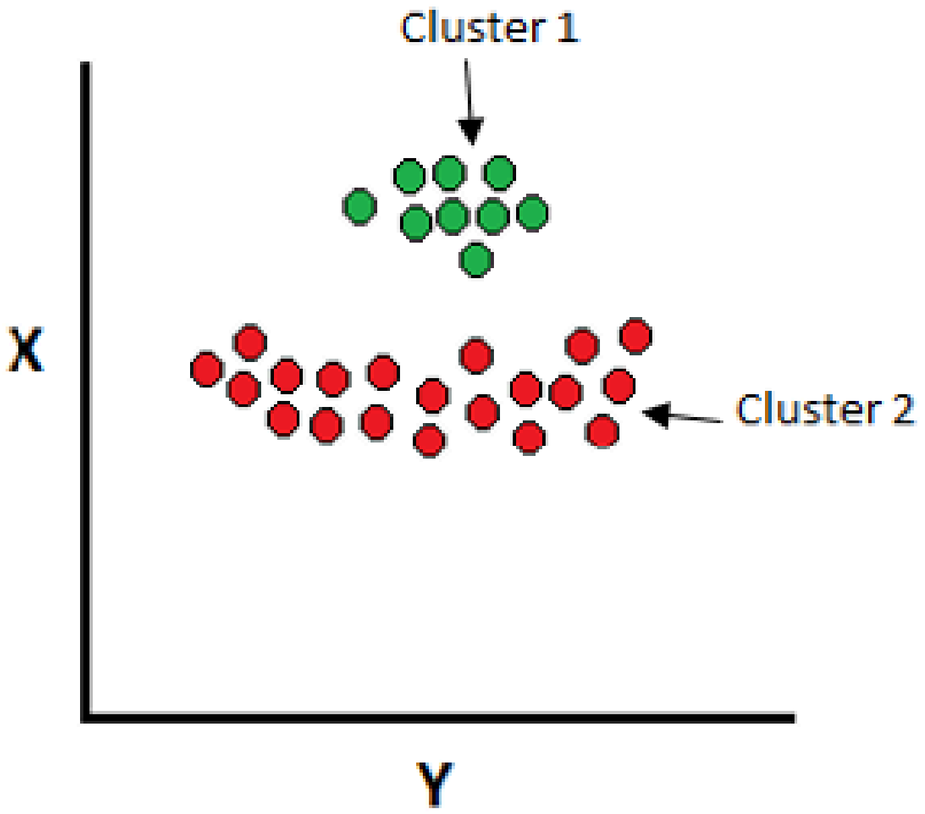
Logistic Regression




Before k-means
After k-means

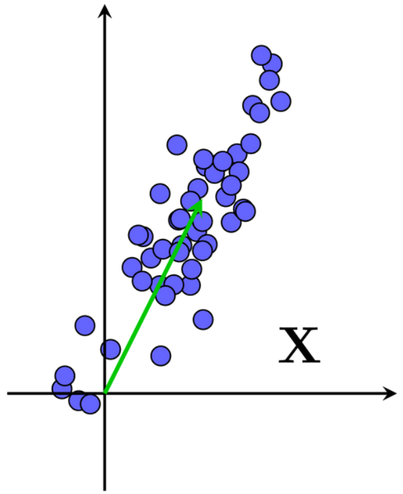
DBScan Clustering



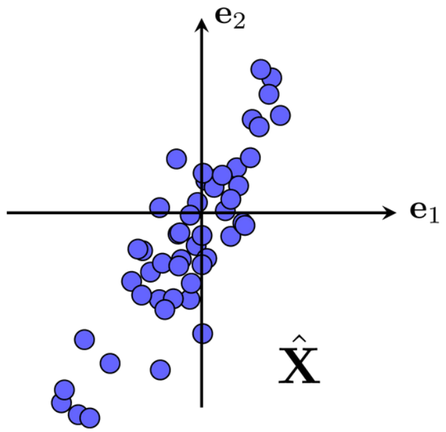


PCA procedure

1. Find mean vector



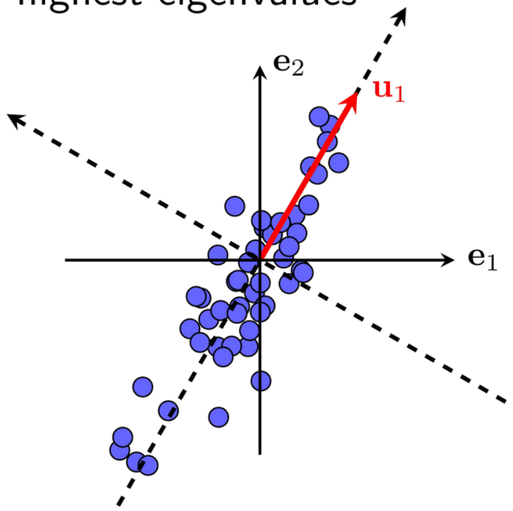
2. Subtract mean



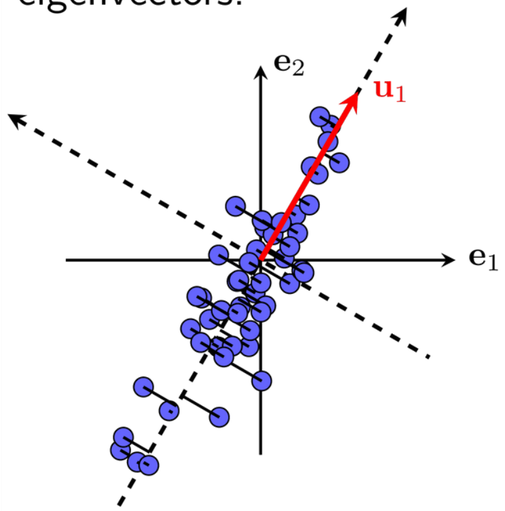
3. Compute covariance matrix:
$\mathbf{S} = \frac{1}{N}\hat{\mathbf{X}}\hat{\mathbf{X}}^T$

4. Computer eigenvalues and eigenvectors of $\mathbf{S}$:
$(\lambda_1, \mathbf{u}_1), \ldots, (\lambda_D, \mathbf{u}_D)$
Remember the orthonormality of $\mathbf{u}_i$.

5. Pick K eigenvectors w. highest eigenvalues



6. Project data to selected eigenvectors.



7. Obtain projected points in low dimension.