KNTU_IPM Machine learning Quiz

Clustering Planetary Orbits Using K-Means

You are an astronomer studying a dataset of 200 exoplanets, each characterized by two physical properties: their orbital period (in days) and planetary radius (in Earth radii).

These properties are derived from observations made by a space telescope, similar to the Kepler mission. Your goal is to use the K-Means clustering algorithm to group these exoplanets into clusters based on their orbital periods and radii, identifying patterns that might correspond to different types of exoplanets, such as hot Jupiters, super-Earths, or mini-Neptunes, without prior labeling.

Dataset Description

- **Data Format:** Each exoplanet is represented as a 2D point [P, R], where:
    - P: Orbital period (days, Preprocessed and normalized to for clustering).
    - R: Planetary radius (Earth radii, Preprocessed and normalized to for clustering).
- **Dataset Size:** 200 exoplanets, each with one [P, R] pair.
- **Objective:** Cluster the exoplanets into K groups (where K is determined using the elbow method) to identify distinct planet types.

Physical Context

- **Hot Jupiters:** Large radii (~10–20 Earth radii), short periods (~1–10 days).
- **Super-Earths:** Moderate radii (~1–2 Earth radii), moderate periods (~10–100 days).
- **Mini-Neptunes:** Moderate radii (~2–4 Earth radii), longer periods (~50–300 days).

Task

1. Implement K-Means clustering using the find_closest_centroids and compute_centroids functions from your KMeans.ipynb notebook.
2. Use the elbow method to determine the optimal number of clusters (K).
3. Visualize the clusters as a scatter plot of orbital period vs. radius, with points colored by cluster.