UNIVERSIDAD OBERTA DE CATALUNYA

Master's degree in Data Science

MASTER THESIS

AREA: Specialized fields

Unsupervised methods applied to Galaxy distribution

Author: Carlos Toro Peñas

Supervisors: ...

Advisors: ...

Professor: Laura Ruíz Dern

REALIA TECHNOLOGIES, SL   August, 2025

Contents

Chapter 1

Introduction

1.1 Justification of intererest and relevance

In this work we want to partially answer question of how density-based algorithms can be applying to figure out how visible matter(1) is shaped arround our position in the Universe.

Now a days we all are furtunated to have public Surveys such as SDSS (Sloan Digital Sky Survey) and 2DFGRS (Two-degree Field Galaxy Redshidft survey) which producing data accesible for everyone. These data contain among others thigs, hundreds of rows representing galaxies with their redshifts. By applying Universe-models like Labmda-CDM to these data it is possible to estimate the distance of such objects to Earth.

How dense the matter is? how galaxies are distributed along the Universe? Is it possible to answer these questions using density-based algorithms?

Density-based unsupervised methods should be able to find the same structure: galaxy clusters integrated on superclusters.

1.2 Personal motivation

As a Master Degree in Data Science I am interested in knowing if density-based algorithms can guess how clusters and superclusters of galaxies are arranged within the Universe.

1.3 Objectives

- Guess how matter is located in a big scale arround Via Lactea.
- Apply densiy algorithm to locate clusters and superclusters of galaxies.