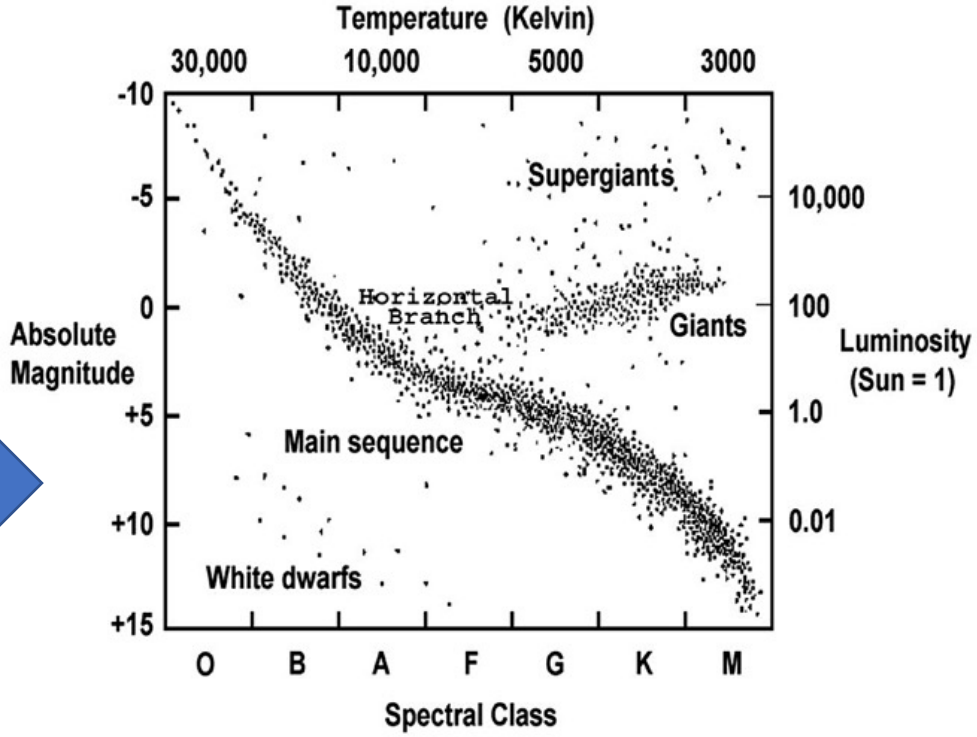
天文数据的结构

问题

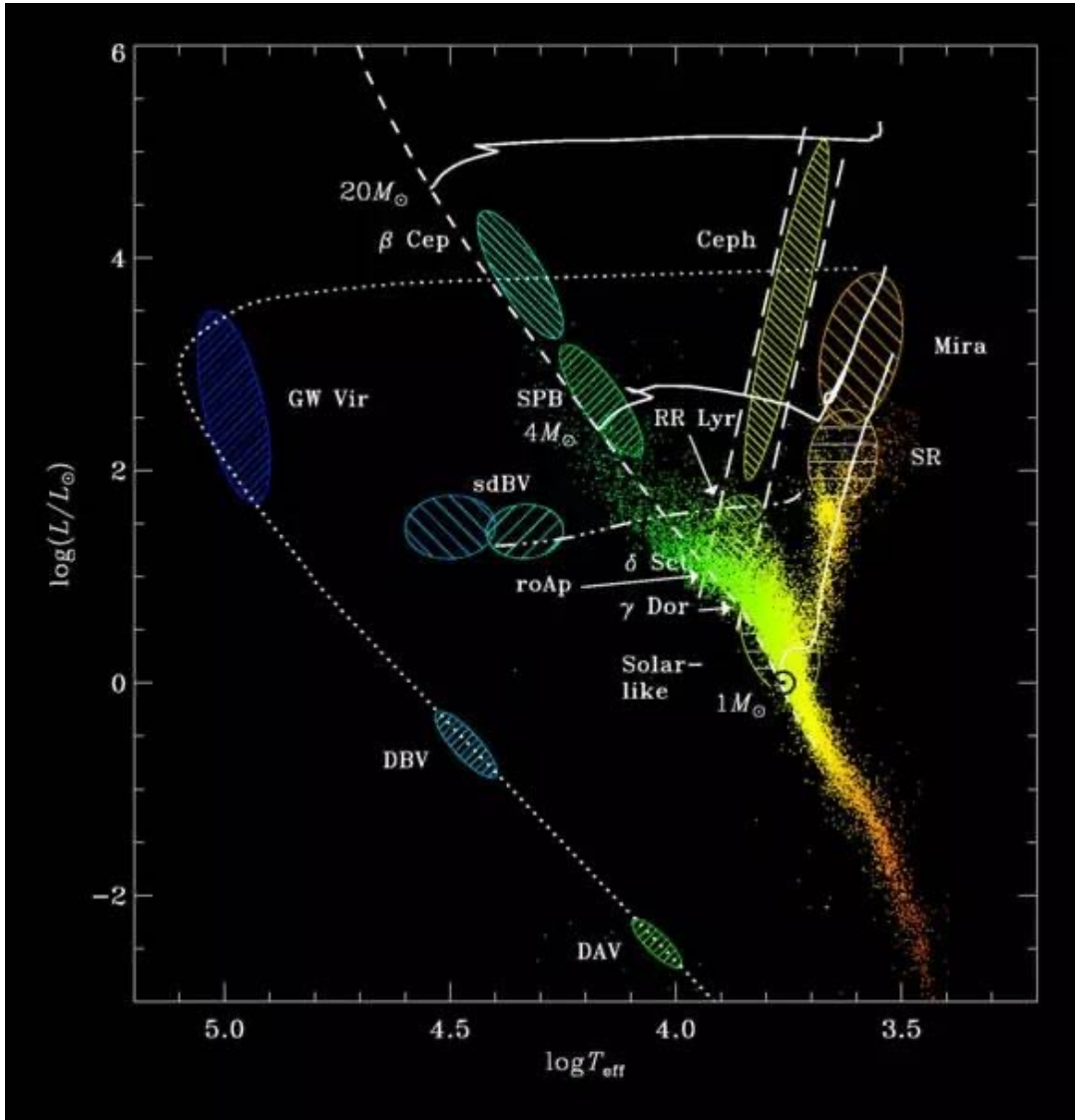
- 我们可以从数据中直接得到它们的结构吗？

1 obsid	2 designation	3 obsdate	4 lmjd	5 mjd	6 planid	7 spid
312121	'J053159.11+3...	2011-10-24	55859	55858	'F5909'	12
509133	'J052930.91+2...	2011-10-24	55859	55858	'M5904'	9
512121	'J053159.11+3...	2011-10-24	55859	55858	'M5904'	12
1901097	'J063100.28+2...	2011-11-08	55874	55873	'GAC_097...	1
3210010	'J054748.37+2...	2011-11-10	55876	55875	'GAC_089...	10
6802178	'J062402.64+2...	2011-11-20	55886	55885	'GAC_0630...	2
7208017	'J013159.85+2...	2011-11-21	55887	55886	'B8704'	8
8002059	'J070408.66+2...	2011-11-24	55890	55889	'GAC_106...	2
8309184	'J032855.00+0...	2011-11-26	55892	55891	'F9204'	9
11101221	'J055540.32+2...	2011-12-03	55899	55898	'GAC_083...	1
14216016	'J105430.43+3...	2011-12-11	55907	55906	'B90706'	16
15207171	'J013159.85+2...	2011-12-13	55909	55908	'M31_019...	7
15213089	'J013637.01+3...	2011-12-13	55909	55908	'M31_019...	13
15508077	'J052832.68+2...	2011-12-14	55910	55909	'GAC_078...	8
15512121	'J053159.11+3...	2011-12-14	55910	55909	'GAC_078...	12
15515066	'J051814.32+2...	2011-12-14	55910	55909	'GAC_078...	15
18214130	'J075853.03+1...	2011-12-19	55915	55914	'F5591504'	14
18309215	'J105656.98+4...	2011-12-19	55915	55914	'F5591506'	9
19104116	'J094002.56+2...	2011-12-21	55917	55916	'F5591705'	4
19704168	'J052658.99+2...	2011-12-22	55918	55917	'GAC_082...	4
19803165	'J013737.21+3...	2011-12-22	55918	55917	'M31_025...	3

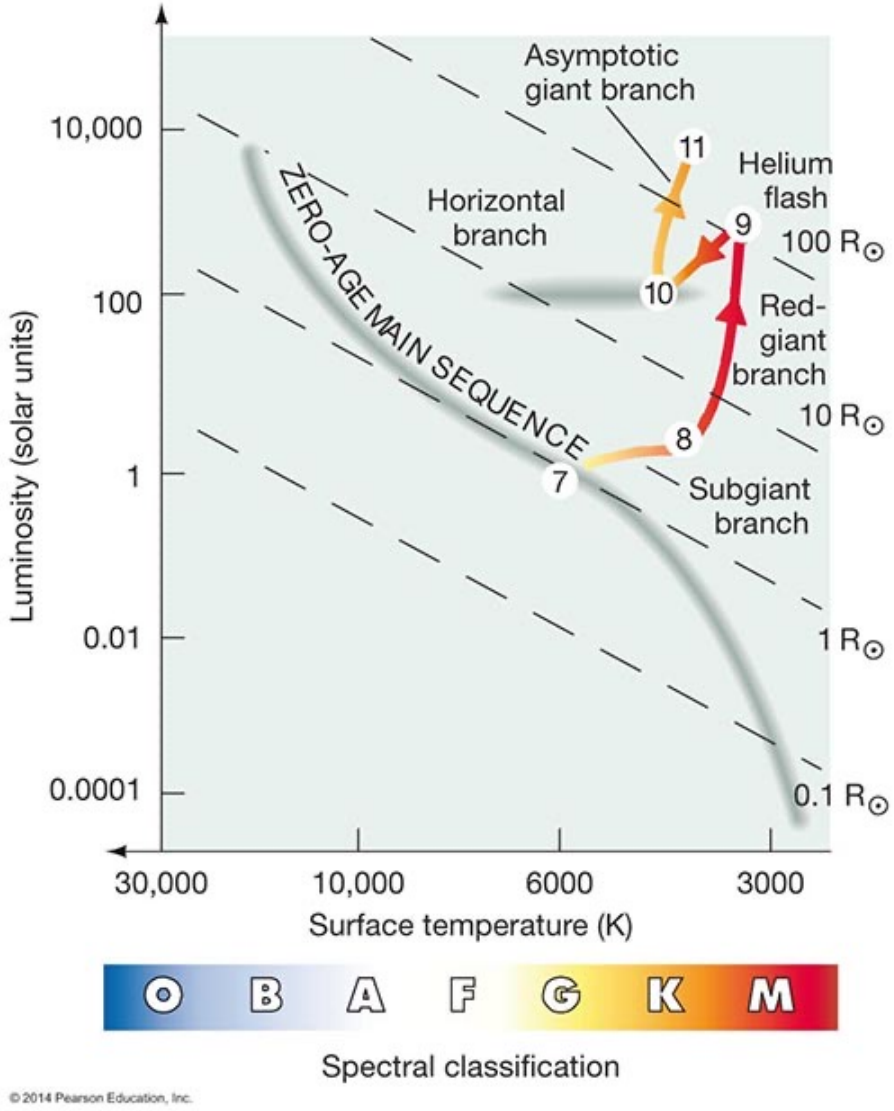


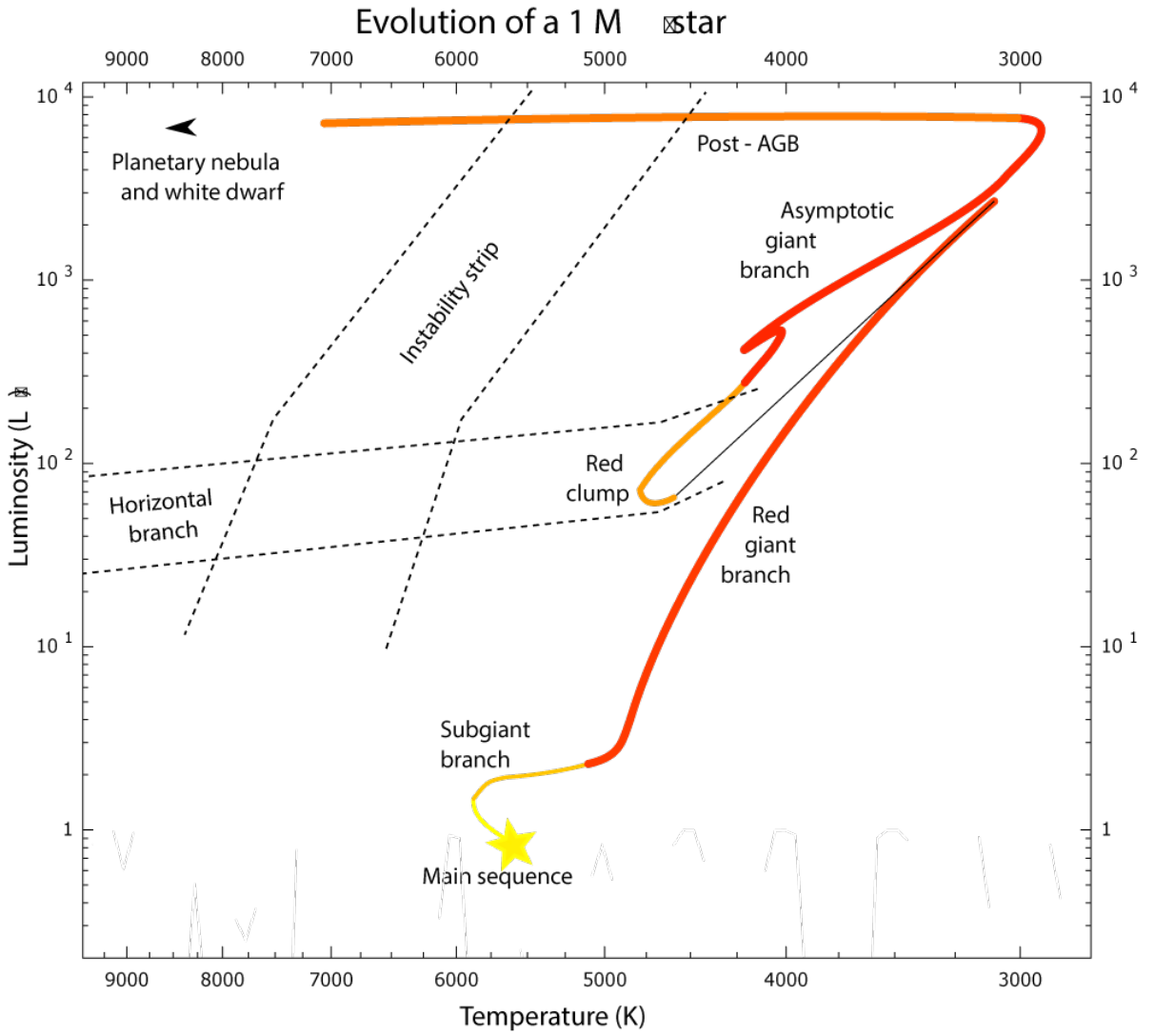
赫罗图 Hertzsprung–Russell diagram

- 来源
  - 主观的人为统计
- 形式
  - log-log图
  - 横坐标为光谱型/温度/色指数（B-V）
  - 纵坐标为光度/绝对星等
- 作用
  - 用于研究恒星的演化



赫罗图


Asymptotic giant branch
Helium flash
Horizontal branch
ZERO-AGE MAIN SEQUENCE
Red-giant branch
Subgiant branch
11
9
10
8
7
10,000
100
1
0.01
0.0001
100 R⊙
10 R⊙
1 R⊙
0.1 R⊙
Luminosity (solar units)
30,000
10,000
6000
3000
Surface temperature (K)
O
B
A
F
G
K
M
Spectral classification
© 2014 Pearson Education, Inc.


Evolution of a 1 M ⊠star
9000
8000
7000
6000
5000
4000
3000
Planetary nebula and white dwarf
Post - AGB
Asymptotic giant branch
Instability strip
Red clump
Horizontal branch
Red giant branch
Subgiant branch
Main sequence
Luminosity (L ⊠)
Temperature (K)

无监督学习 Unsupervised machine learning

- 没有人为输入标签
- 多用于形态学分类
- 少部分用于分析聚类与离群点

1. arXiv:2206.06787 [pdf, other] astro-ph.IM
galmask: A Python package for unsupervised galaxy masking
Authors: Yash Gondhalekar, Rafael S. de Souza, Ana L. Chies-Santos
Abstract: Galaxy morphological classification is a fundamental aspect of galaxy formation and evolution studies. Various machine learning tools have been developed for automated pipeline analysis of large-scale surveys, enabling a fast search for objects of interest. However, crowded regions in the image may pose a challenge as they can lead to bias in the learning algorithm. In this Research Note, we prese… ▿ More
Submitted 11 June, 2022; **originally announced** June 2022.
Comments: Submitted to RNAAS

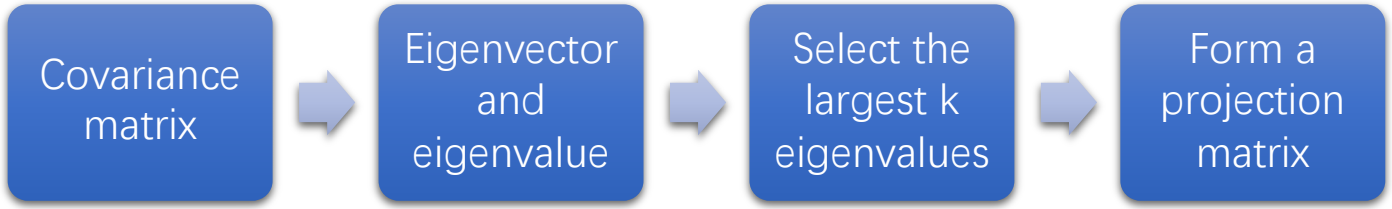
2. arXiv:2206.06165 [pdf, ps, other] cs.LG astro-ph.GA astro-ph.IM
The Classification of Optical Galaxy Morphology Using Unsupervised Learning Techniques
Authors: Ezra Fielding, Clement N. Nyirenda, Mattia Vaccari
Abstract: The advent of large scale, data intensive astronomical surveys has caused the viability of human-based galaxy morphology classification methods to come into question. Put simply, too much astronomical data is being produced for scientists to visually label. Attempts have been made to crowd-source this work by recruiting volunteers from the general public. However, even these efforts will soon fail… ▿ More
Submitted 13 June, 2022; **originally announced** June 2022.
Comments: 6 pages, to be presented at the International Conference on Electrical, Computer and Energy Technologies (ICECET 2022) 20-22 July 2022, Prague-Czech Republic

3. arXiv:2205.09760 [pdf, other] cs.CV astro-ph.IM
Identifying outliers in astronomical images with unsupervised machine learning
Authors: Yang Han, Zhiqiang Zou, Nan Li, Yanli Chen
Abstract: Astronomical outliers, such as unusual, rare or unknown types of astronomical objects or phenomena, constantly lead to the discovery of genuinely unforeseen knowledge in astronomy. More unpredictable outliers will be uncovered in principle with the increment of the coverage and quality of upcoming survey data. However, it is a severe challenge to mine rare and unexpected targets from enormous data… ▿ More
Submitted 19 May, 2022; **originally announced** May 2022.

4. arXiv:2205.09344 [pdf, other] astro-ph.GA astro-ph.CO astro-ph.IM
Unsupervised classification of CIGALE galaxy spectra
Authors: J Dubois, D Fraix-Burnet, J Moultaka, P Sharma, D Burgarella
Abstract: Aims. The present study aims at providing a deeper insight into the power and limitation of an unsupervised classification algorithm (called Fisher-EM) on spectra of galaxies. This algorithm uses a Gaussian mixture in a discriminative latent subspace. To this end, we investigate the capacity of this algorithm to segregate the physical parameters used to generate mock spectra and the influence of t… ▿ More
Submitted 19 May, 2022; **originally announced** May 2022.
Comments: Astronomy and Astrophysics - A&A, EDP Sciences, In press

5. arXiv:2205.01129 [pdf, other] astro-ph.GA hep-ph
Measuring Galactic Dark Matter through Unsupervised Machine Learning
Authors: Matthew R Buckley, Sung Hak Lim, Eric Putney, David Shih
Abstract: Measuring the density profile of dark matter in the Solar neighborhood has important implications for both dark matter theory and experiment. In this work, we apply autoregressive flows to stars from a realistic simulation of a Milky Way-type galaxy to learn -- in an unsupervised way -- the stellar phase space density and its derivatives. With these as inputs, and under the assumption of dynamic e… ▿ More
Submitted 2 May, 2022; **originally announced** May 2022.
Comments: 23 pages, 9 figures

PCA



- $f_{gas,neutral} - M_{stellar} - SFR$
  - $f_{gas,neutral}$: neutral gas fraction, $f_{gas,neutral} = \frac{M_{HI}+M_{H2}}{M_{HI}+M_{H2}+M_{stellar}}$
  - $M_{stellar}$: stellar mass
  - SFR: star formation rate

- Sample: Evolution and Assembly of GaLaxies and their Environments (EAGLE)
  - $M_{stellar} > 10^9 M_\odot$
  - SFR>$0.01 M_\odot\ yr^{-1}$
  - $M_{neutral} > 10^7 M_\odot$
  - $0 \leq z \leq 4.5$

The Fundamental Plane of star formation in galaxies revealed by the EAGLE hydrodynamical simulations FREE

Claudia del P. Lagos ✉, Tom Theuns, Joop Schaye, Michelle Furlong, Richard G. Bower, Matthieu Schaller, Robert A. Crain, James W. Trayford, Jorryt Matthee

Monthly Notices of the Royal Astronomical Society, Volume 459, Issue 3, 01 July 2016, Pages 2632–2650, https://doi.org/10.1093/mnras/stw717

Published: 01 April 2016 **Article history ▼**

TDA

- 检测星际介质间的可压缩磁流体动力湍流
- 采用GASS数据，观测中性H密度进行分析
- 并未涉及大样本数据
- 有明确的目的

Topological data analysis and diagnostics of compressible MHD turbulence

I. Makarenko[1]†, P. Bushby[1], A. Fletcher[1], R. Henderson[1], N. Makarenko[2] and A. Shukurov[1]

[1]School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[2]Central Astronomical Observatory, Russian Academy of Sciences, 65 Pulkovskoye Chaussee, Saint-Petersburg, 196140, Russia

(Received xx; revised xx; accepted xx)

有关TDA的愿景

大样本数据集

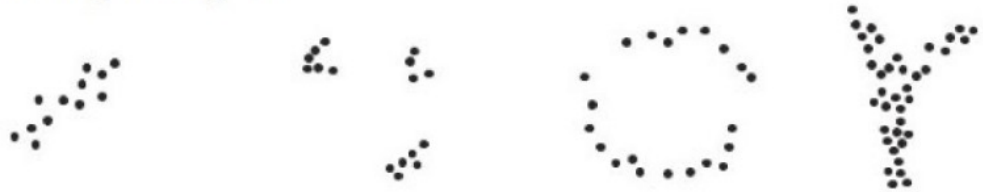
↓

无监督的给出数据集结构

↓

结构背后的物理过程

Data sets come in many shapes and sizes. The data sets presented in the figure below illustrate this point very well.



- The data set on the far left has the rough shape of a line in the plane. We are all familiar with many examples of this kind of data, and such data are typically handled with various regression models, which permit prediction and also allow for greater understanding of the data. This helps in developing mental models.
- The second set from the left illustrates a data set which decomposes into disjoint groups, and is not well approximated by any line. This kind of data occurs very frequently in the biomedical and social sciences, and cluster analysis has been developed to produce such decompositions in order to deliver taxonomies for the data.
- The third set from the left is a type of data set that occurs frequently when one is dealing with time series data representing periodic or recurrent behavior of some kind.
- The data set on the far right might describe data in which there are one standard or normal mode and three extremal modes. For example, it might come from sensors on an airliner, where the standard mode is flying at altitude in non-turbulent conditions and where the three extremal modes are takeoff, landing, and flying at altitude in turbulent conditions.