

Introduction to Clustering

What we will be covering in this module?

- Supervised vs Unsupervised Learning
- Clusters and Clustering
- Properties of Clusters
- Applications of Clustering
- K-means
- Hierarchical Clustering

Supervised Learning

Supervised Learning

Regression

Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
Medium	Tier 1	Supermarket Type1	3735.1380
Medium	Tier 3	Supermarket Type2	443.4228
Medium	Tier 1	Supermarket Type1	2097.2700
NaN	Tier 3	Grocery Store	732.3800
High	Tier 3	Supermarket Type1	994.7052

Supervised Learning

Regression

Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
Medium	Tier 1	Supermarket Type1	3735.1380
Medium	Tier 3	Supermarket Type2	443.4228
Medium	Tier 1	Supermarket Type1	2097.2700
NaN	Tier 3	Grocery Store	732.3800
High	Tier 3	Supermarket Type1	994.7052

Classification

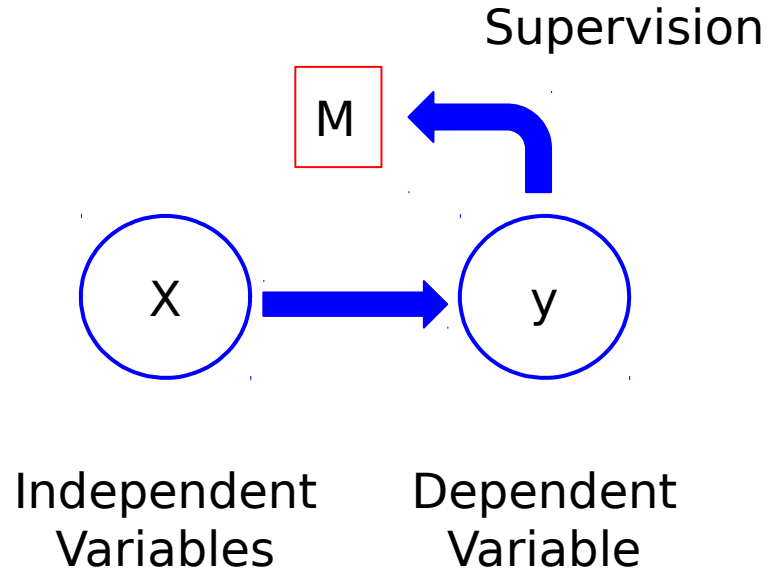
Pclass	Sex	Age	Survived
3	male	22.0	0
1	female	38.0	1
3	female	26.0	1
1	female	35.0	1
3	male	35.0	0

Supervised Learning

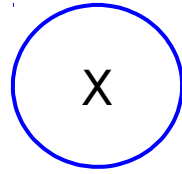
X			y
Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
Medium	Tier 1	Supermarket Type1	3735.1380
Medium	Tier 3	Supermarket Type2	443.4228
Medium	Tier 1	Supermarket Type1	2097.2700
NaN	Tier 3	Grocery Store	732.3800
High	Tier 3	Supermarket Type1	994.7052

X			y
Pclass	Sex	Age	Survived
3	male	22.0	0
1	female	38.0	1
3	female	26.0	1
1	female	35.0	1
3	male	35.0	0

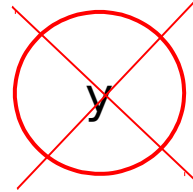
Supervised Learning



Unsupervised Learning

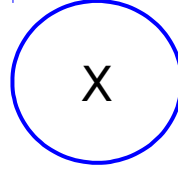


Independent
Variables

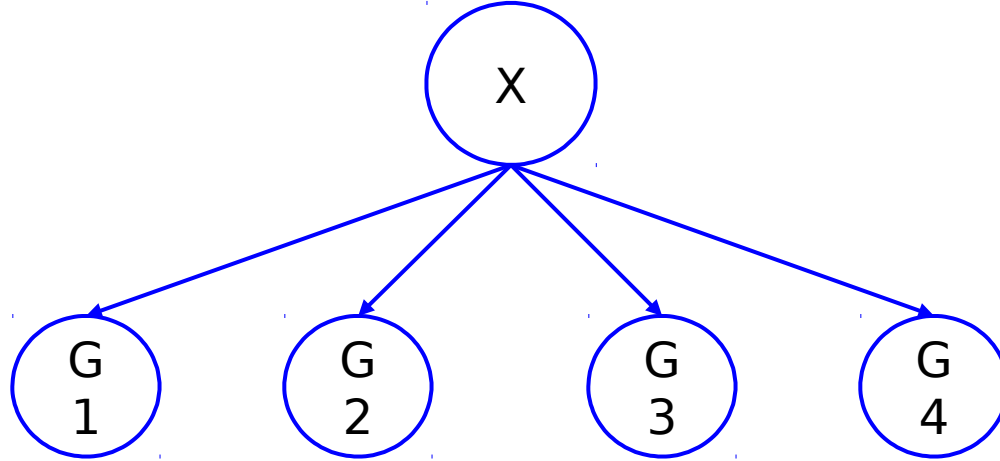


Dependent
Variable

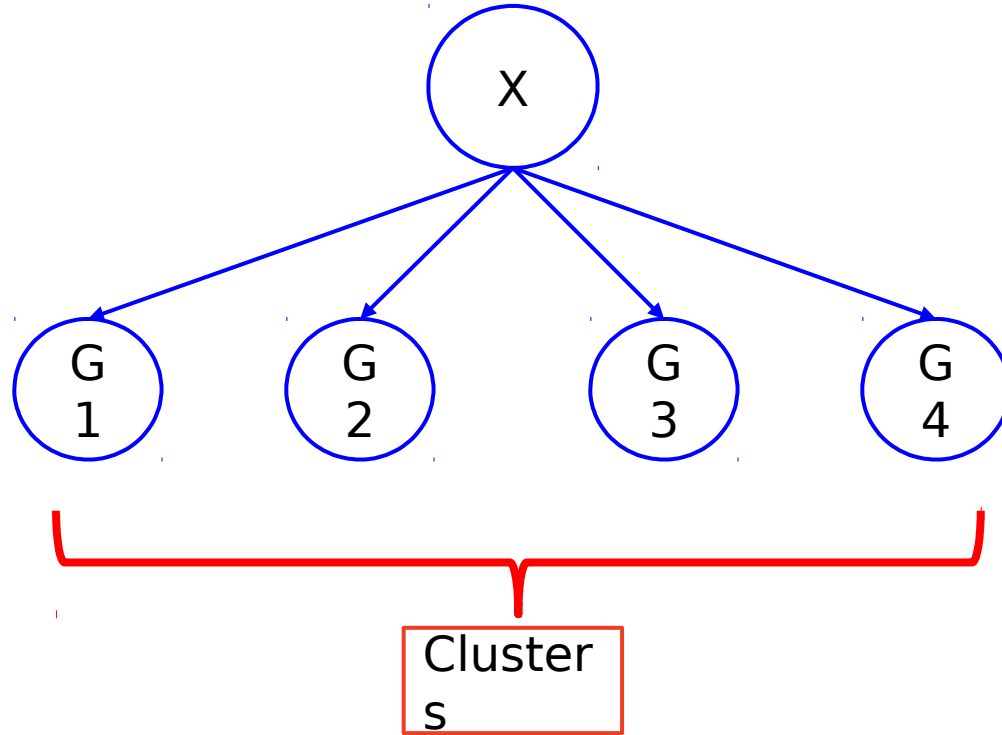
Unsupervised Learning



Unsupervised Learning



Clustering



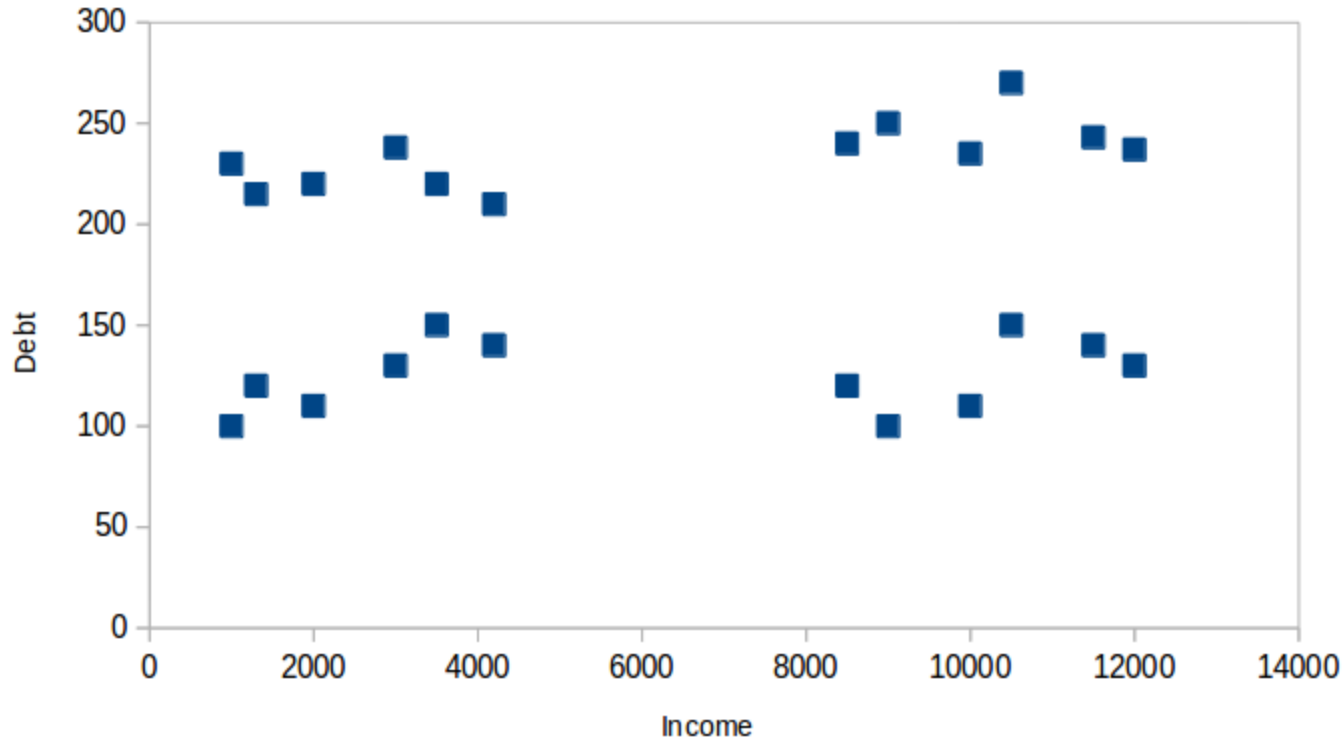
Customer Attributes

- Income
- Occupation
- Gender
- Debt
- Age
- Credit Score
- City of residence
- Marital Status

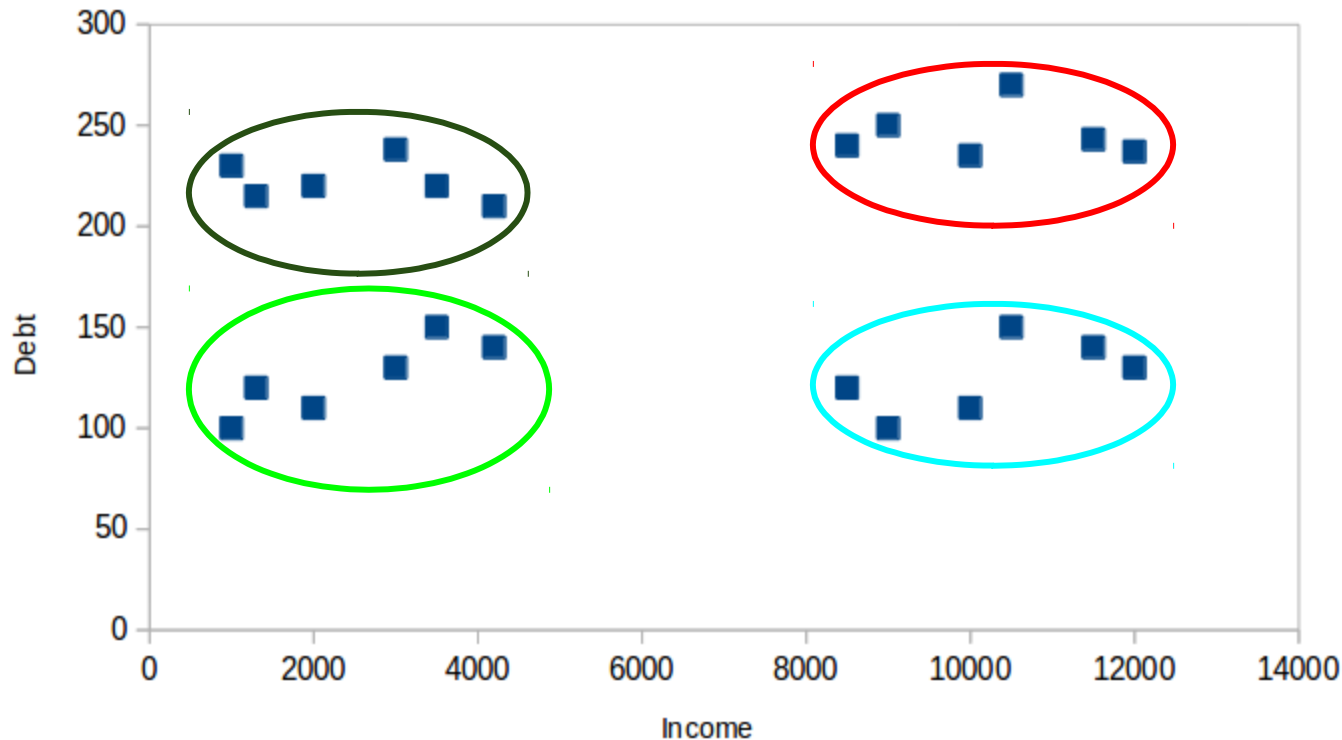
Customer Attributes

- Income
- Debt

Customer Attributes



Customer Segmentation



Properties of Clusters

Properties of Clusters

- Data points within same cluster should be similar

Properties of Clusters

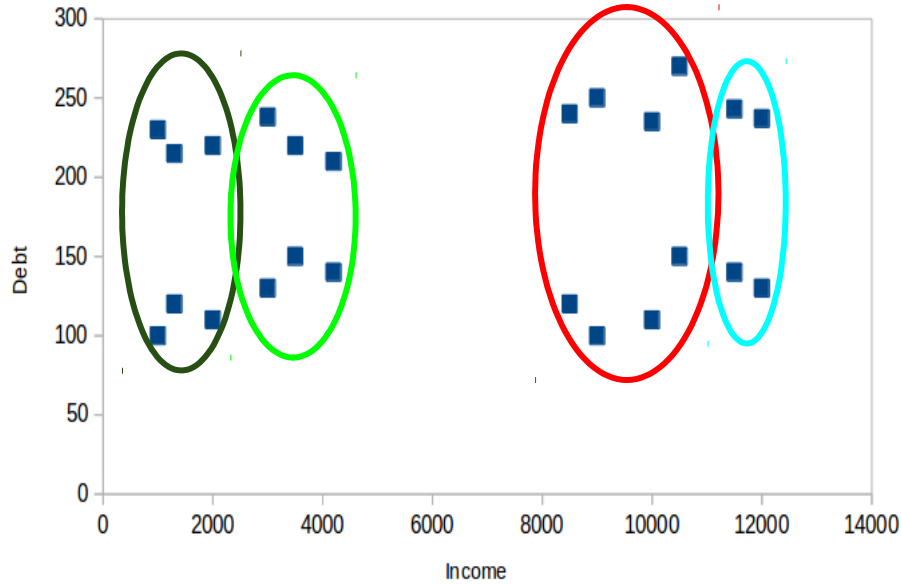
- Data points within same cluster should be similar



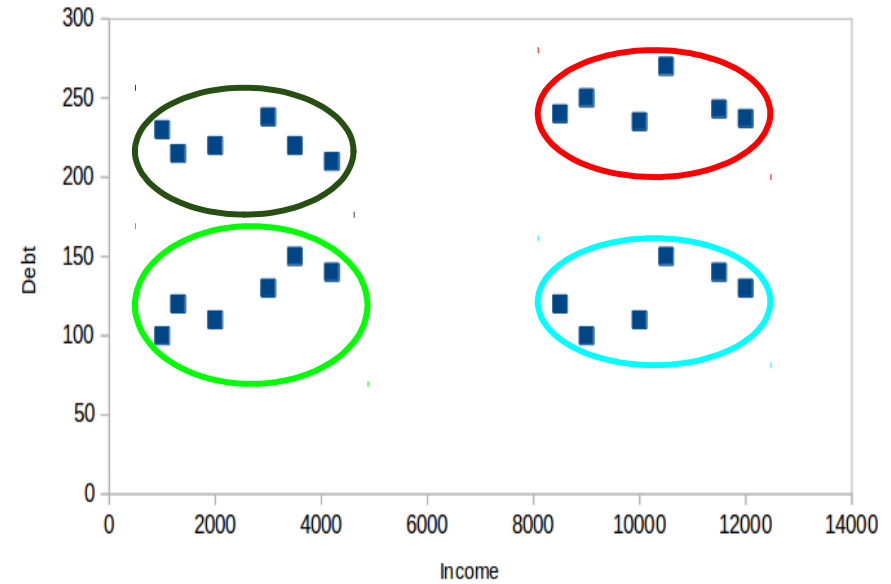
Properties of Clusters

- Data points within same cluster should be similar.
- Data points from different clusters should be as different as possible.

Properties of Clusters



Case - I



Case - II

Properties of Clusters

- Data points within same cluster should be similar.
- Data points from different clusters should be as different as possible.

Thank
You!