

Avaliação 2 - Análise Multivariada 2

Tharso Monteiro

Análise de Agrupamento e Análise Discriminante

7 de dezembro de 2021

Para aplicação da análise de agrupamento, foi utilizado o conjunto de dados *Credit Card Dataset for Clustering*, disponível em <https://www.kaggle.com/arjunbhasin2013/ccdata>. Este conjunto de dados contém informações de 9000 usuários ativos de cartões de crédito durante os últimos seis meses antes da coleta dos dados. As variáveis contidas no conjunto descrevem, a frequência de compras e empréstimos feitos com o cartão de débito e crédito, limite, saldo, frequência de atualizações do saldo, entre outras informações.

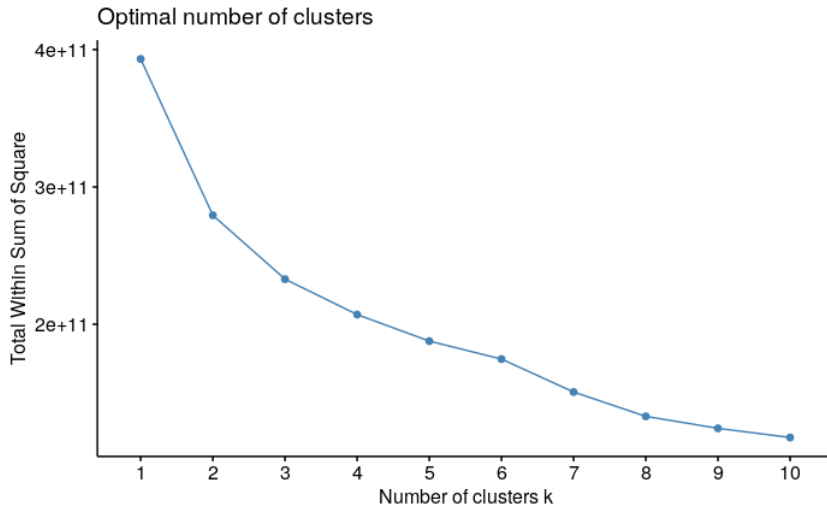


Figura: WCSS por número de *clusters*

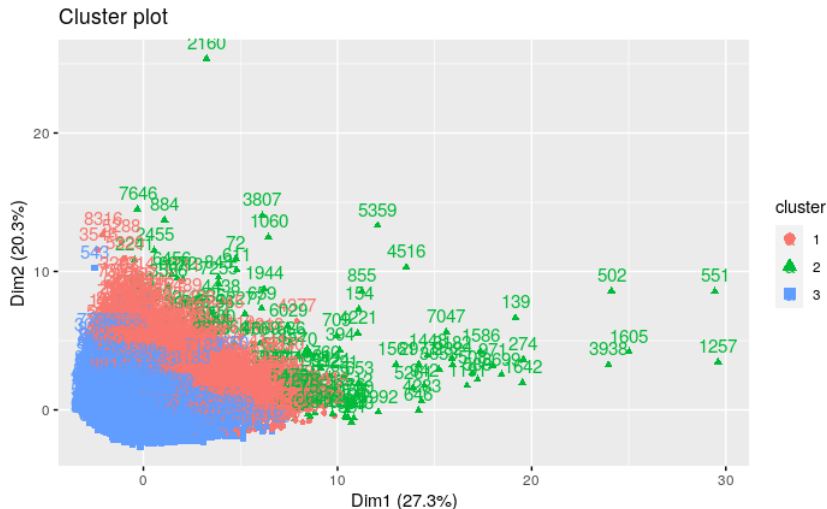


Figura: Agrupamento de acordo com os 2 primeiros CPs

Analizando os componentes principais, vemos que o primeiro componente atribui peso maior às variáveis referentes ao número e frequência de compras, enquanto o segundo atribui peso maior ao valor e frequência de empréstimos.

Deste modo, podemos atribuir a seguinte interpretação aos agrupamentos :

- Agrupamento 1 : Usuários não-frequentes do cartão de crédito.
- Agrupamento 2 : Usuários moderados do cartão de crédito.
- Agrupamento 3 : Usuários severos do cartão de crédito.

Para a aplicação do agrupamento hierárquico, utilizamos o conjunto de dados *Economic Data and Statistics on World Economy and Economic Freedom*, da fundação Heritage.

Esse conjunto de dados contém informações sobre indicadores econômicos e índices ligados à liberdade econômica em 173 países.

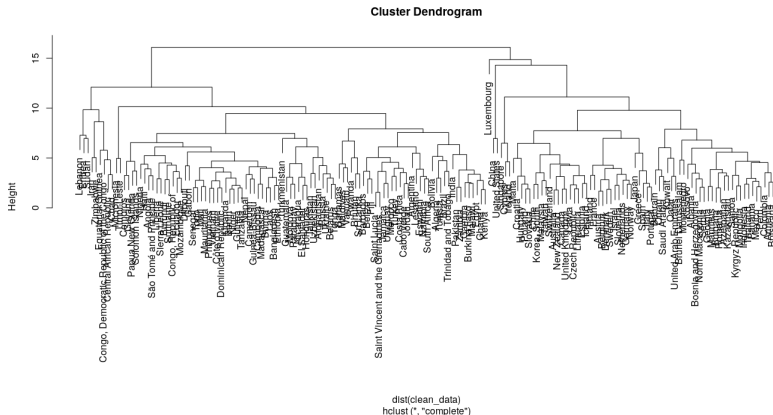


Figura: Dendrograma da análise de agrupamento hierárquica com distância máxima

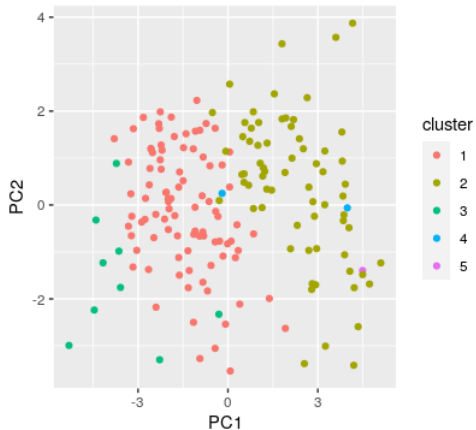


Figura: Agrupamento por componentes principais

O primeiro componente principal atribui peso positivo maior às variáveis que indicam maior liberdade financeira, de negócios, investimentos e maior PIB *per capita*. O segundo componente principal atribui pesos negativos em alto valor absoluto para as variáveis referentes aos impostos. Isto é, países com valores maiores da segunda componente principal têm taxas de imposto mais baixas.

- Cluster 1 : Países subdesenvolvidos e em desenvolvimento. Neste grupo se encontram a maior parte dos países do sudeste asiático, América Latina e África, assim como a Índia e Rússia. Este é o agrupamento mais numeroso.
- Cluster 2 : Países desenvolvidos que não são potências globais. Nesse grupo se encontram a maioria dos países europeus, Austrália, Nova Zelândia, e países da Ásia e América Latina com maiores índices de liberdade econômica, como Taiwan, Japão, Panamá, Chile, México e Peru.
- Cluster 3 : Países em conflito ou em grave crise econômica. Nesse grupo se encontram : República do Chade, Eritreia, Irã, Sudão, Zimbábue, Líbano, entre outros.
- Cluster 4 : Potências mundiais. Nesse agrupamento, se encontram somente a China e os Estados Unidos.
- Cluster 5 : Neste agrupamento, se encontra apenas Luxemburgo, um pequeno país europeu com altíssimos níveis de liberdade econômica.

Para a demonstração do agrupamento baseado em misturas, utilizamos o conjunto de dados **banknote**, contido no pacote **mclust**. Esse conjunto contém dados de seis medições feitas em 100 notas verdadeiras e 100 notas falsas de mil francos suíços.

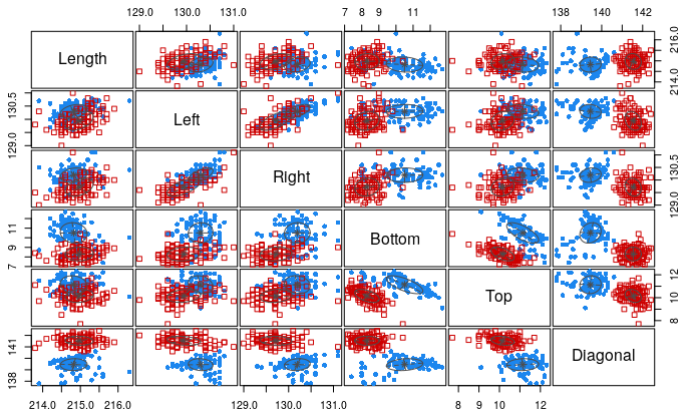


Figura: Agrupamento por pares de variáveis no modelo de misturas

	Agrupamento 1	Agrupamento 2
Falsificação	100	0
Genuíno	1	99

Tabela: Matriz de confusão do agrupamento por misturas

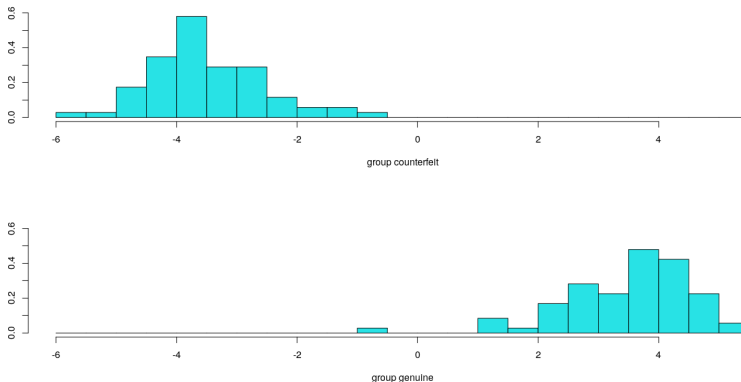


Figura: Histograma por grupo : Análise Discriminante Linear

	Agrupamento 1	Agrupamento 2
Falsificação	31	0
Genuíno	0	29