### **Módulo 7 - Aprendizagem não supervisionada** Aula 52a - Introdução ao clustering hierárquico

**Objetivos de aprendizagem**

1. Introduzir conceitos vinculados ao clustering hierárquico
2. Entender o que é um dendrograma
3. Avançar na compreensão de aspectos vinculados às diferentes medidas de diferença
4. Identificar as diferenças e semelhanças com o método de K-Means

**Conteúdo**

1. Clustering hierárquico
2. Dendrogramas: interpretação
3. Medidas de diferença (“Linkage”):
   1. Single linkage
   2. Complete linkage
   3. Average linkage
   4. Ward linkage
4. Diferenças e semelhanças com K-Means

**Modalidades de aula incluídas**

|  |  |
| --- | --- |
| hablar.png | Exposição teórica oral e interativa, com material de apoio por escrito |

|  |  |
| --- | --- |
|  | Práticas interativas, grupais e individuais, usando notebooks |

### Roteiro da aula

|  |  |
| --- | --- |
| 1. **Boas-vindas** | **5 minutos** |

|  |  |
| --- | --- |
| hablar.png | * Verificar se os PCs estão funcionando corretamente * Fazer uma revisão das noções gerais de clustering k-means e sua avaliação |

|  |  |
| --- | --- |
| 1. **Clustering hierárquico: Conceito** | **10 minutos** |

|  |  |
| --- | --- |
| hablar.png | * Visão geral do algoritmo * Visualização gráfica |

|  |  |
| --- | --- |
| 1. **Interpretação do dendrograma** | **15 minutos** |

|  |  |
| --- | --- |
| hablar.png | * Importância visual do dendrograma:   + permite ter uma ideia visual da quantidade de clusters ideal observada no conjunto de dados * Destacar a disposição gráfica do dendrograma:   + Ao final do dendrograma, cada observação é apenas um cluster.   + Conforme vamos subindo pela árvore, algumas folhas começam a se juntar: aquelas que correspondem a observações muito semelhantes.   + Conforme vamos subindo pela árvore, um número cada vez maior de observações se juntou. Quanto mais cedo (parte mais baixa da árvore) duas observações se juntam, mais similares são entre si.   + As observações que se juntam mais para cima são as mais diferentes. |

|  |  |
| --- | --- |
| 1. **Formação de clusters em clustering hierárquico: Linkage** | **25 minutos** |

|  |  |
| --- | --- |
|  | * Mostrar as formas mais comuns e as diferenças entre elas   + Single linkage   + Complete linkage   + Average linkage   + Ward linkage * Apresentar o Copehenetic Correlation Coefficient para medir a correlação entre as distâncias entre clusters e as distâncias entre pontos originais * Destacar como o uso de diferentes métodos de linkage dá resultados diferentes |

|  |  |
| --- | --- |
| 1. **Prática guiada: Clustering hierárquico no conjunto de dados IRIS** | **30 minutos** |

|  |  |
| --- | --- |
|  | * Utilizando SciPy, construir um dendrograma e avaliar a quantidade de clusters que existem a partir do método de linkage de Ward * Fazer uma inspeção visual deles * Calcular outros dendrogramas para outros métodos de linkage (complete, single, average e weighted) e comparar os resultados |

|  |  |
| --- | --- |
| 1. **Conclusão** | **5 minutos** |

|  |  |
| --- | --- |
| hablar.png | * Apresentar métodos de clustering hierárquicos * Oferecem uma vantagem em relação a K-Means, pois não exigem a definição prévia da quantidade de clusters a criar. * Oferecem uma representação gráfica (dendrograma) do processo de geração de clusters * Existem vários métodos para definir a forma de união (linkage) entre os diferentes clusters |