Gravação da aula 1 de DCBD: https://drive.google.com/file/d/1JeITad9l3piFYBlKPTgq47s1vobbDVMO/view?usp=sharing

Gravação da aula 2 de DCBD: https://drive.google.com/file/d/1OWCreeBwmyYhAzA7U8RcjQxX1pPkk0\_l/view?usp=sharing

Gravação da aula 3 de DCBD: https://drive.google.com/file/d/1zsNAZCxRdxqAYg4MGc6fKHD1cz8wlLeu/view?usp=sharing

Gravação da aula 4 de DCBD: https://drive.google.com/file/d/1nOTh1ermUKe0i\_SmyDmGKNw2MJbiNdsR/view?usp=sharing

1. **2 formas de ocasionar overfitting (tenta fazer uma linha perfeita que aborda todas as amostras):**

- poucas amostras disponíveis

- ruídos na imagem

1. **ligação tomek**

- fronteira de decisão

- 2 exemplos de classes diferentes muito próximas (mais próximos do q entre a mesma classe)

- remove exemplos que formam a ligação tomek

- as vezes pode remover as classes de fronteira, influenciando no resultado do algoritmo

1. **KNN**

- os exemplos mais importantes são os que estão próximos a fronteira de decisão

- o KNN calcula a distância e remove os mais longes da fronteira de decisão

- limpeza de vizinhança (também é bom para eliminar ruídos)

1. **Seleção de características mais importantes para o domínio do problema**

- 1º passo é a seleção manual, feita pelo especialista do domínio

**- construir ou remover atributos**

**- Wrapper = força bruta**

1. **Mineração dos dados**
2. **Regressão = utiliza o passado para predizer o futuro**
3. **Predição = prevê o futuro, utilizando dados do passado. Ex covid 19, valores da ação, pela ordem temporal do problema**
4. **Agrupamento = quem tá próximo fica junto. É necessário um especialista humano pra ver os resultados. Ex. Fraldas e cervejas no supermercado. Quem tem valores próximos fica no mesmo grupo**
5. **Última fase: Avaliação dos dados**