



Aula 07 - Publicando o Portfólio

🕒 Created	@January 5, 2022 6:55 PM
📎 Materials	
☑ Reviewed	<input type="checkbox"/>

Introdução: Pensamento Analítico de Dados:

-

Parte 01: (Processamento de Dados e "Big Data" - Dados e Capacidade de Data Science como um Ativo Estratégico)

- “Eu tenho preocupação em entender o quão o espaço separa as classes”.
- Histogramas:
 - Eu tenho uma visualização sobre como o espaço separa as classes
 - Histograma é a projeção dos pontos em 1 dimensão. Me ajude a representar o espaço e entender o quão separável.
 - “Se eu utilizar somente 1 atributo para separar as classes, haverá uma mistura”.
 - A inferência fala muito sobre os erros da mistura.
 - “Eu adapto um conjunto de hiperplanos separadores para realizar recortes no espaço e criar regiões de separação”
- Histogramas:
 - Eixo x => todos os valores.
 - Eixo y => o número de ocorrências daquele determinado valor em x
 - O número de ocorrências representa a densidade dos valores.

- “Existe sempre uma transformação espacial que permite a construção de um espaço onde eu posso usar funções mais simples. Eu consigo reescrever meu problema para um formato mais simples”.
- Gerar projeções e realizar estudos algébricos para verificar o espaço multidimensional.
- Correlação entre variáveis:
 - Mede-se ao grau do relacionamento linear entre variáveis.
 - Qual é a dependência linear que existe entre essas duas variáveis?
 - Correlação igual a 1 significa que a variável 1 é exatamente igual a 2 seguindo uma tendência linear.
 - Correlação igual a 0 significa que a variável 1 não tem nenhuma relação linear com a variável 2.
 - Alta correlação entre duas variáveis significa que você pode escolher uma variável ou outra, porque elas tem características muito similares.
- Correlação de Pearson
 - Trabalha em cima dos dados
 - Covariância: Produto interno entre duas variáveis centralizadas. O produto interno permite que eu verifique similaridade.
 - Da álgebra linear, a projeção permite com que eu meça quanto de similaridade existe entre dois vetores em um determinado espaço.
 - Similaridade positiva significa que os vetores formam um ângulo agudo menor que 90 graus
 - Similaridade negativa significa que os vetores formam um ângulo obtuso maior que 90 graus.
 - O produto interno permite a medição de similaridade entre duas funções. O resultado diz o quão importante a similaridade é.
 - $x - \text{mean}(x) \cdot y - \text{mean}(y) / (\text{sd}(x) \cdot \text{sd}(y)) / \text{número de pontos} \Rightarrow$ similaridade média por ponto.
 - Probabilidade vem do conceito de uma medida.
- Correlação de Spearman
 - Trabalha em cima da “rank” dos dados

- Rank é a posição relativa do elemento em termos de ordem.
- Exemplo:
 - Dados: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0
 - Rank: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
- Há cenários onde o que importa é a relação dos valores que as variáveis tem entre si, enquanto que em outros cenários, o que importa é a ordem relativa que os valores das variáveis tem entre si.
- Há cenários que o importante são as grandezas.

Histograma:

Correlação de Pearson entre variáveis:

- Definições:
 - Mede a dependência linear entre variáveis.
 - Relação entre 2 variáveis seguindo uma reta.
 - “Essa variável é igual a outra, seguindo uma tendência linear”.

Produto Interno

- Mede a similaridade entre 2 vetores.
- A projeção permite que eu meça a similaridade entre 2 vetores.
- O valor resultante do produto interno diz a importância da correlação.
- Se dividir o resultado do produto interno pelo número de pontos, obtemos a similaridade média por ponto.