

Relatório LAMIA - 5

Prática: Estatística p/ Aprendizado de Máquina (I)

Kaique Medeiros Lima

Introdução

Este relatório foca em apresentar o conteúdo de estatística aprendido no curso 'Machine Learning, Data Science and Generative AI with Python'. Esses conceitos foram aplicados em Python, apresentando também o Matplotlib, uma biblioteca de visualização de dados, pois disponibiliza a criação de vários tipos de gráficos.

Descrição da atividade

Tipos de dados:

- **Numéricos:** Quantitativos, podem ser medidos. Podem ser contínuos ou discretos, contínuos podem possuir casas decimais, discretos são valores contáveis, não podendo ser fracionáveis.
- **Catégoricos:** Representa categorias ou grupos, não realiza operações aritméticas.
- **Ordinais:** Possuem hierarquia, como bom, médio e ruim.

Média: Soma de valores divididos pela sua quantidade. Oferece um número para comparações.

Mediana: Em um vetor ordenado de números, pega o número localizado no meio. Se for um valor par, soma os dois centrais e divide por 2.

Moda: Número que aparece com mais frequência no vetor.

Variância: Indica o quão dispersos estão os valores em um vetor. Valores baixos de variância indicam um conjunto homogêneo, valores altos indicam heterogeneidade.

Desvio Padrão: Indica o grau de dispersão ou variabilidade dos valores em relação à média. É a raiz quadrada da variância.

População: Conjunto completo de todos os elementos.

Amostra: Parte selecionada da população.

Histograma: É um gráfico que representa a distribuição dos dados em barras.

Distribuição normal/Gaussiana: Representa a probabilidade de valores em torno de uma média, com a maioria dos dados perto da média e diminuindo conforme se afastam. É simétrica e tem a forma de um sino.

Distribuição Exponencial / PDF: Representa o tempo entre eventos em um processo de Poisson, onde eventos ocorrem de forma contínua e

independente em uma taxa constante. Normalmente é uma curva decrescente, mostrando que a probabilidade de esperar um evento cresce com o tempo.

Distribuição Binomial / BMF: Descreve o número de sucessos em um número fixo de tentativas, onde cada tentativa pode ser apenas um fracasso ou sucesso.

Distribuição de Poisson / PMF: Número de eventos acontecendo em um intervalo fixo de tentativas, esses eventos ocorrem com uma taxa média constante e de forma independente.

Percentil: O percentil informa qual porcentagem dos dados está abaixo de um determinado valor.

Momentos: O primeiro momento é a média, o segundo a variância, o terceiro momento mede a assimetria da distribuição, o quarto é a curtose da função.

Assimetria: Indica se os dados estão distribuídos de maneira simétrica perto da média.

Curtose: Descreve a forma das pontas de uma distribuição em relação a uma distribuição normal.

Covariância: Indica o grau de variação entre duas variáveis. Se a covariância for positiva, as variáveis aumentam ou diminuem juntas; se for negativa, uma variável aumenta enquanto a outra diminui.

Correlação: Quantifica a força e a direção da relação linear entre duas variáveis. Varia de -1 a 1, onde 1 indica uma correlação positiva perfeita, -1 indica uma correlação negativa perfeita, e 0 indica nenhuma correlação.

Probabilidade Condicional: Probabilidade de um evento acontecer, dado que outro evento já ocorreu primeiro e que se depende.

$$P(B | A) = (P(A, B)) / (P(A))$$

Teorema de Bayes: Descreve a probabilidade de um evento acontecer baseado no conhecimento de condições necessárias para o sucesso do evento.

$$P(A | B) = (P(A) * P(B | A)) / (P(B))$$

Crash Course de matplotlib e Seaborn:

O vídeo apresenta também um curso rápido para as bibliotecas Matplotlib e Seaborn, ambas servem para a criação de diversos tipos de gráficos. Matplotlib produz gráficos mais básicos, já o Seaborn, oferece gráficos mais prazerosos ao olhar.

Conclusão

Esse card foi focado no reforço dos conceitos fundamentais básicos de estatística, também trabalhado em Python. Também foi apresentado duas novas bibliotecas, o Matplotlib e o Seaborn, que são essenciais no dia a dia do

aprendizado de máquinas para ter melhor entendimento de como o código está funcionando por meio de gráficos detalhados, o Seaborn oferece um modo agradável de criar gráficos.