

Prática: Estatística p/ Aprendizado de Máquina

Eduardo Prasniewski

O curso assistido serviu imensamente para relembrar alguns conceitos de estatística e probabilidade, matéria qual já cursei anteriormente, porém fazia pouco uso do que foi aprendido de forma prática. Como durante esta matéria e de Cálculo Numérico eu também usei jupyter e numpy para realizar as atividades, boa parte do que foi apresentado também já tinha conhecimento, segue então as minhas anotações tomadas durante as videoaulas:

Tipos de dados:

- Numérico: Pode ser discreto ou contínuo.
- Categórico: Dado qualitativo.
- Ordinal: Uma junção de numérico e categórico (um dado numérico que também é qualitativo)

O professor passa de forma prática a aplicação e uso das seguintes definições, usando a distribuição normal da estatística.

- Média = Soma/quantidade
- Mediana = Numero que se encontra ao meio da amostra.
- Moda = Valor que mais aparece.
- Desvio padrao = Índice de uniformidade dos dados
- Variança = Também uma medida de dispersão.

Mostra algumas formas de distrubuição, como de Poisson, a Normal ou Gaussiana e a Binomial de Newton. Apresenta visualização de percentuais e os momentos, na seguida ordem: media, variação, skew (a torção da distribuição) e kurtosis (quão acentuada é a distribuição).

Aprensenta um mini-curso de matplotlib, de forma prática, a plotagem de gráficos e seus variações, como histogramas, scatters, pizza, boxplot etc. Da mesma forma introduz o seaborn, um superset feito em cima da matplotlib, tem um design mais moderno e possui algumas features a mais, como em um mesmo gráfico haver o histograma e a interpolação dos dados.

Por fim introduz mais uma vez de forma prática os conceitos de covariância e correlação entre duas variáveis. Logo após, demonstra a base da probabilidade condicional, algo que é relativamente complicado de entender de começo os conceitos, mas depois de dar uma pesquisada entendi direito. Por último, apresenta o teorema de Bayes e explica um caso de uso dele, os resultados de testes de uso de drogas, onde mesmo tendo uma alta precisão, tem uma probabilidade alta de gerar também falsos negativos/positivos.