



Identificação de outliers multivariados através de técnicas estatísticas e computacionais

ANTONIO FREDERICO SOARES BRUM BRUMANO (Autor), TIAGO MARTINS PEREIRA (Orientador)

Em qualquer conjunto de dados, as observações que apresentam um grande afastamento das restantes ou são inconsistentes com elas são habitualmente designadas por outliers. Isso ocorre por erro no armazenamento dos dados ou por existirem realmente alguns pontos diferentes dos demais. Os outliers podem ser univariados (uma variável) ou multivariados (duas ou mais variáveis); no caso univariado o outlier pode ser detectado analisando-se o gráfico de dispersão que mostra o comportamento de cada observação do conjunto de dados; no caso multivariado utilizando gráfico, a detecção torna-se mais complexa, uma vez que seria preciso analisar duas variáveis por vez, já que alguns pontos poderiam ser atípicos com relação a algumas variáveis e não ser em relação a outras. A identificação de outliers é importante em muitas aplicações reais, tais como detecção de fraudes, diagnóstico de falhas, monitoramento de condições médicas, dentre outras. Muitas das primeiras propostas para a identificação de outliers referem-se a métodos baseados na análise gráfica. Basicamente estes métodos referem-se ao uso do box-plot e por meio de gráficos de dispersão. Para um conjunto de dados multivariado, a detecção de outliers pode ser feita através de métodos estatísticos ou métodos baseados em mineração de dados e aprendizado de máquinas.

Instituição de Ensino: Universidade Federal de Ouro Preto