

ANÁLISE DE COMPONENTES PRINCIPAIS E EQUIDADE

Felipe Tetsuo Yamada

Faculdade de Ciências Aplicadas - FCA
R. Pedro Zaccaria, 1300, Limeira - SP, 13484-350
felipeyamada@gmail.com

Leonardo Tomazeli Duarte

Faculdade de Ciências Aplicadas - FCA
R. Pedro Zaccaria, 1300, Limeira - SP, 13484-350
leonardo.duarte@fca.unicamp.br

RESUMO

Um cenário frequente na aplicação de métodos de aprendizado de máquina é a manipulação de grandes volumes de dados. O treinamento de uma rede neural, por exemplo, requer o processamento de muita informação, tendo como consequência um alto custo computacional. Dessa forma, uma prática comum na etapa de pré-processamento de dados é a redução de dimensionalidade por meio de técnicas como a Análise de Componentes Principais (Principal Component Analysis, PCA, em inglês). O método tem como entrada uma matriz de covariância e, como saída, os autovetores e autovalores da respectiva matriz, permitindo a projeção dos dados em dimensões reduzidas e o controle da quantidade da informação que será perdida no processo. Devido à sua formulação matemática, PCA pode acabar favorecendo uma das duas classes dos dados a serem analisados. O projeto pretende avaliar o desempenho do método PCA em relação a uma versão modificada do método, chamada Fair PCA, tendo como base uma população de dados com apenas duas classes distintas. O novo método traz resultados considerados mais equilibrados que o método tradicional de PCA e permite um "trade-off" de perdas, visando favorecer a classe mais prejudicada e desfavorecer a classe privilegiada com o intuito de gerar resultados menos discrepantes. Há também interesse na análise de desempenho de outras técnicas, como: Fatoração De Matrizes Não-negativas (NMF), Side-channel Analysis (SCA) e Independent Component Analysis (ICA).

PALAVRAS CHAVE. Viés. ACP. Pré-processamento de dados. Aprendizado Não-Supervisionado.

AM - Aprendizado de Máquina, PSD - Programação Semidefinida