

**MBA
USP
ESALQ**

Unsupervised Machine Learning: Exercícios Aplicados

Prof. Dr. Wilson Tarantin Junior

*A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Introdução

- Técnicas exploratórias (não supervisionadas)
 - Para avaliar a relação conjunta entre as variáveis (interdependência)
 - Não há modelos do tipo “ $y_i = x_{1i} + x_{2i} + \dots + u_i$ ”
 - Não são adequadas para fins de inferência, ou seja, não têm caráter preditivo para observações de fora da amostra
 - Se novas observações forem adicionadas ao banco de dados, é adequado refazer a análise; o mesmo é válido se variáveis forem adicionadas ou retiradas

Objetivo – Análise de Cluster

- Agrupamento de observações do banco de dados
 - O objetivo é agrupar as observações em grupos homogêneos internamente e heterogêneos entre si
 - **Dentro do grupo:** observações semelhantes com base nas variáveis utilizadas na análise
 - **Entre grupos distintos:** observações diferentes com base nas variáveis utilizadas na análise

Objetivo – Análise Fatorial PCA

- Agrupamento das variáveis **métricas** em fatores
 - O objetivo é o agrupamento das variáveis em fatores
 - A técnica depende das correlações entre variáveis
- Algumas possíveis finalidades:
 - Obter o comportamento conjunto de variáveis, combinando-as para redução estrutural
 - Análise da validade de construtos pela identificação das variáveis alocadas aos fatores
 - Elaboração de rankings para classificação de desempenho por meio dos fatores
 - Criação de fatores ortogonais entre eles e posterior uso em modelos supervisionados

Objetivo – Anacor e ACM

- Análise de associação entre variáveis **categóricas**
 - O objetivo é verificar se existe associação estatisticamente significativa entre as variáveis qualitativas (duas ou mais de duas variáveis)
 - Como estão associadas as categorias das variáveis?
 - Busca-se criar o mapa perceptual para visualizar tais associações
 - Qual a proximidade das categorias no mapa perceptual?

Referência

Fávero, Luiz Paulo; Belfiore, Patrícia. (2017). Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. Rio de Janeiro: Elsevier