MAE5776 – Análise Multivariada 1º Sem/2022 Questões -Consolidando o Aprendizado

Aluno:

Thiago Ferreira Miranda - nº USP: 11925711

- 1. Entender a estrutura dos dados sob análise é importante em todas as análises estatísticas? Particularmente, em análise multivariada de dados, discuta por que isso é importante.
- 2. Como a teoria de Espaços Duais pode ser útil na redução de dimensionalidade de dados?
- 3. Justifique a afirmação: "As técnicas matriciais de Decomposição em Valores Singulares, bem como de Decomposição Espectral, são a base de muitas das análises de redução de dimensionalidade".
- 4. Considere a matriz de dados Y_{nxp} e as correspondentes formas quadráticas YY' e Y'Y.
- (a) Se λ é um autovalor de Y'Y com autovetor v. Mostre que λ é um autovalor de YY' com autovetor Yv (equivalentemente, com autovetor padronizado $Yv\lambda^{-1/2}$).
- (b) Como esse resultado pode ser usado para relacionar Componentes Principais e Coordenadas Principais?
- 5. Considere a matriz de dados Y_{nxp} e as correspondentes formas quadráticas YY' e Y'Y.
- (a) Estabeleça relações entre os autovalores e autovetores da decomposição em valores singulares da matriz retangular Y e das correspondentes decomposições espectrais das formas quadráticas
- (b) Como esse resultado pode ser usado em big-data?
- 6. O Biplot é uma ferramenta útil na visualização de dados multivariados. Explique por que e como esse gráfico é construído.
- 7. Como a Análise de Correspondência pode ser formulada a partir da Análise de Coordenadas Principais?
- 8. Na Análise de Fatores (Análise Fatorial), como estão definidos os fatores comuns e os específicos? Você pode usar a solução via Componentes Principais para responder.
- 9. Na análise de redução de dimensionalidade de uma matriz de dados Y_{nxp} , os Componentes Principais satisfazem quais propriedades? O que garante que dois componentes reduzem bem os dados?

- 10. Na análise de redução de dimensionalidade de uma matriz de dados Y_{nxp} , os Eixos Discriminantes da Solução Linear de Fisher satisfazem quais propriedades? O que garante que dois eixos discriminantes reduzem bem os dados?
- 11. Na análise de redução de dimensionalidade de uma matriz de dados Y_{nxp} , com $p = p_1 + p_2$, os Eixos Canônicos da Correlação Canônica satisfazem quais propriedades? O que garante que o primeiro par desses eixos reduzem bem os dados?
- 12. Considere a redução de dimensionalidade de uma matriz de dados Y_{nxp} . Se n << p, quais são os problemas na realização da Análise de Componentes Principais "Clássica"? Que alternativas de análise podem ser usadas?
- 13. Na MANOVA qual é a importância da equação envolvendo o seguinte determinante: $|S_B \lambda S_W| = 0$, em que S_B e S_W são matrizes quadradas (pxp), conhecidas, de soma de quadrados e produtos cruzados dos efeitos Entre e Dentro de grupos, respectivamente, e $\lambda \mathbb{R}^+$ é tal que $(S_W^{-1}S_B)V = \lambda V$, com $V \in \mathbb{R}^p$?
- 14. Na análise de uma matriz de dados multivariados Y_{nxp} , como a distância de Mahalanobis pode ser usada para definir Regiões de Concentração dos dados e Regiões de Confiança para o centroide? Como pode ser proposto um critério de diagnóstico de observações atípicas (ouliers)?
- 15. Na análise de uma matriz de dados multivariados Y_{nxp} (,) com p = 2, ilustre, em um gráfico de dispersão, possíveis diferenças entre os Intervalos de Confiança Univariados Clássicos, Intervalos de Confiança Univariados Simultâneos, Intervalos de Confiança Univariados com Correção de Bonferroni e Regiões de Confiança para o vetor μ . Compare essas quatro abordagens no contexto de correções para múltiplos testes.
- 16. Em uma análise estatística, considere que na normalização de dados foi usada a seguinte transformação: $y_{ij}-v_{ij}z_{1i}$

em que, y_{ij} é a resposta do i-ésimo paciente na j-ésima variável, z_{1i} é o escore desse paciente no primeiro componente principal e v_{ij} é o j-ésima coordenada do correspondente autovetor.

- (a) Apresente situações práticas do por que adotar tal transformação dos dados em uma análise estatística.
- (b) Apresente críticas ao uso dessa estratégia de normalização de dados.
- (c) Em problemas gerais da análise de dados, cite alternativas que podem ser usadas para normalização de dados.
- 17. Apresente uma situação prática para ilustrar a utilidade da Análise de Componentes Principais e de Coordenadas Principais.
- 18. Apresente uma situação prática para ilustrar a utilidade da Análise de Correspondência.
- 19. Apresente uma situação prática para ilustrar a utilidade da Análise Fatorial (fatores Comuns e específicos).
- 20. Apresente uma situação prática para ilustrar a utilidade da Análise Discriminante e da MANOVA.

- 21. Apresente uma situação prática para ilustrar a utilidade da Análise de Correlação Canônica.
- 22. Com o apoio da Aula em que geramos dados e visualizamos direções dos vetores reducionistas de dados multivariados, compare as soluções de Componentes Principais, Análise Discriminante, Correlação Canônica e Regressão Linear.
- 23. Qual é o problema analítico e possíveis soluções da análise de dados em Big-n?