Lista 2 Multivariada

Davi Wentrick Feijó

2023-05-02

8) Provar o seguinte teorema: Sejam A e B matrizes idempotentes. Então,

(a) $A + B$ é idempotente somente quando $AB = BA = 0$.

$$(A + B)^2 = A + B$$
$$A^2 + AB + BA + B^2 = A + B$$

Como A e B são idempotentes, temos $A^2 = A$ e $B^2 = B$.

$$A + AB + BA + B = A + B$$
$$AB + BA = A + B - B - A$$
$$AB + BA = 0$$

(b) $C = AB$ é idempotente somente quando $AB = BA$.

$$C^2 = ABAB = AB$$

Agora, queremos manipular essa equação para obter AB = BA. Para isso, podemos reescrever ABAB da seguinte forma

$$ABAB = AB$$

Se AB = BA podemos substituir na equação

$$(AB)(BA) = AB$$
$$A(BB)A = ABA = AB$$

Se AB = BA podemos substituir novamente na equação.Sabendo que A e B sao idempotentes!

$$(AB)A = (BA)A = B(AA) = BA = AB$$

Como A e B sao idempotentes, temos:

$$A$$

Para essa equacao ser igual a zero temos duas opceos:

$$AB = 0 \qquad AB - I = 0 \Rightarrow AB = I$$

(c) $I - A$ é idempotente.

Temos que mostrar que:

$$(I - A)^2 = I - A$$

Vamos abrir o quadrado

$$I^2 - IA - AI + A^2 = I - A$$

$$I - 2A + A^2 = I - A$$

Como A é idempotente entao $A^2 = A$

$$I - 2A + A = I - A$$
$$I - A = I - A$$

Acabmos de mostrar que

$$(I - A)^2 = (I - A)$$

9) Provar o seguinte teorema:

Seja $X(n \times x)$ tal que $rank(X) = k < n$. Então, $P_X = X(X^tX)^{-1}X^t$ é idempotente e simétrica e consequentemente, uma matriz projeção ortogonal.

Vamos mostrat que é idempotente, para isso temos que mostrar que $(P_X)^2 = P_X$

$$(P_X)^2 = [X(X^tX)^{-1}X^t][X(X^tX)^{-1}X^t]$$
$$(P_X)^2 = X(X^tX)^{-1}X^tX(X^tX)^{-1}X^t$$

$$(P_X)^2 = X(X^tX)^{-1}[X^tX(X^tX)^{-1}]X^t$$
$$(P_X)^2 = X(X^tX)^{-1}IX^t = X(X^tX)^{-1}X^t = (P_X)$$

Agora vamos mostrar que é simetrica. Uma matriz é simétrica se sua transposta é igual a ela mesma. Vamos mostrar que $(P_X)^T = P_X$.

$$(P_X)^T = [X(X^TX)^{-1}X^T]^T$$
$$(P_X)^T = (X^T)[(X^TX)^{-1}]^TX$$

$$(P_X)^T = X^T[(X^TX)^{-1}]^TX$$
$$(P_X)^T = X^T[(X^TX)^T]^{-1}X$$
$$(P_X)^T = X^T[XX^T]^{-1}X$$

$$(P_X)^T = X^T[X^{-1}(X^T)^{-1}]X$$
$$(P_X)^T = X^T[X^{-1}(X^{-1})^T]X$$

10) Utilizando o R: verifique, através de exemplos, que uma matriz de projeção tem autovalores somente no conjunto {0, 1}. A demonstração pode ser feita utilizando a equação característica e lembrando que se M é uma matriz de projeção, então $M = M^2 = M^T$.

11) Seja X uma matriz de dados $(n \times p)$ com matriz de covariância S. Sejam $\lambda_1, ..., \lambda_p$ os autovalores de S.

(a) Mostre que a soma das variâncias s_{ii} de X (variação amostral total) é dada por $\lambda_1 + ... + \lambda_p$

Para provar que a soma das variâncias s_{ii} de X é igual a $\lambda_1 + ... + \lambda_p$, vamos analisar a relação entre os autovalores da matriz de covariância S e as variâncias dos dados.

A matriz de covariância S é definida como a matriz simétrica (p $\times$ p) cujos elementos S_{ij} são dados pela fórmula:

$$S_{ij} = cov(X_i, X_j) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_j - \bar{X})}{n}$$

Onde x_i e x_j são as colunas correspondentes nos dados da matriz X, e $cov(x_i, x_j)$ é a covariância entre essas duas variáveis.

A variância de uma variável x_i é dada pelo elemento S_{ii} da matriz de covariância S. Portanto, podemos escrever:

$$S_{ii} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{n} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = Var(X_i)$$

A soma das variâncias s_{ii} de X é dada por:

$$s_{11} + s_{22} + ... + s_{pp}$$

No entanto, os elementos da matriz de covariância S são os autovalores multiplicados pelos seus correspondentes autovetores. Assim, podemos expressar os elementos da matriz de covariância S como:

Sabemos pela definicao de autovetores e autovalores que eles seguem a seguinte equacao:

$$SA = \lambda A$$

Onde S é a nossa matriz de dados, A é nossa matriz de autovetores de S e λ é nossos autovalores. Em seguida podemos calcular a variancia dos pontos projetados sobre os autovetores A

$$Var(SA) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij}A_j - \mu)^2}{n}$$

Podemos centralizar nossos dados (absorvendo a media para dentro da variavel) deixando com média 0:

$$\mu = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij}A_j)}{n} = \frac{1}{n}\sum_{j=1}^{k}(\sum_{i=1}^{n} X_{ij})A_j = \sum_{j=1}^{k}(\frac{1}{n}\sum_{i=1}^{n} X_{ij})A_j$$

Podemos perceber que se centralizarmos os dados, ou seja subtarair a média, nossa media será zero. Isso acontece pois na equacao que chemaos estamos tirando as medias $S_{.j}$ para cada j o que no caso encontrariamos 0 pois centralizamos os dados diminuindo sua media da coluna.

Podemos continuir densenvolvendo a conta abrindo esse quadrado:

$$Var(SA) = \frac{1}{n}\sum_{i=1}^{n}(\sum_{j=1}^{k}(X_{ij}A_j)\sum_{d=1}^{k}(X_{id}A_d)$$

Vamos chamar cada somatorio com uma letra diferente para nao se confundir(eles sao a mesma coisa!!!)

Podemos reorganizar a soma da seguinte forma, colocando os somatorios com i para dentro junto com o $1/n$ e os que depende de j e a para fora

$$Var(SA) = \sum_{d=1}^{k}\sum_{j=1}^{k}(\frac{1}{n}\sum_{i=1}^{n}X_{id}X_{ij})A_dA_j$$

Pode se perceber que o temos que encontramos entre parenteses nada mais é do que a covariancia entre 2 variaveis que tem média 0:

$$Var(SA) = \sum_{d=1}^{k}(\sum_{j=1}^{k}cov(X_d, X_j)A_j)A_d$$

Vale notar que estamos multiplicando a $Cov(S_d, S_J)$ por um autovetor j da matriz A. Pela relacao $SA = A\lambda$. Sabmos que isso deve ser igual a:

$$\sum_{j=1}^{k}cov(X_d, X_j)A_j = \lambda A_d$$

Agora podemos substituir isso na equcao anterior:

$$Var(SA) = \sum_{d=1}^{k}(\lambda A_d)A_d$$

$$Var(SA) = \sum_{d=1}^{k}\lambda(A_dA_d)$$

Sabemoss que os autovetore sao vetores unitarios logo esse produto interno é igual a 1

$$Var(SA) = \sum_{d=1}^{k}\lambda$$

O que acabamos de fazer é calcular a variancia de XA que é a variancia da projecao dos valores de X sobre a matriz de autovetores A.

Assim, concluímos que a soma das variâncias sii de X (variação amostral total) é igual a $\lambda_1 + ... + \lambda_p$.

(b) Mostre que a variância amostral generalizada é dada por $\lambda_1 \times ... \times \lambda_p$.
(c) Mostre que a variância amostral generalizada se anula se as colunas de X somarem zero.