

## 2. INTRODUÇÃO AOS MÉTODOS FACTORIAIS

### CONCEITOS GEOMÉTRICOS. INÉRCIA.

Os métodos factoriais de Análise de Dados permitem descrever matrizes  $X$  (segundo o modelo do Quadro  $Q$  da Fig. 2.1) de dimensão  $(n, p)$  que representam os valores tomados por  $p$  propriedades em  $n$  indivíduos;  $x_{ij}$  é o valor tomado pela propriedade (variável)  $j$  no indivíduo (amostra)  $i$ .

A matriz  $X$  pode ser tomada segundo as linhas ou segundo as colunas:

- Cada linha de  $X$  é um vector que representa a posição de um indivíduo no espaço das propriedades  $R^p$ .
- Cada coluna de  $X$  é um vector que representa a posição de uma propriedade no espaço dos indivíduos  $R^n$ .

Assim a matriz pode ser assimilada a duas nuvens de pontos distintas, conforme o espaço escolhido, como se pode visualizar no exemplo esquemático da Fig. 2.1.

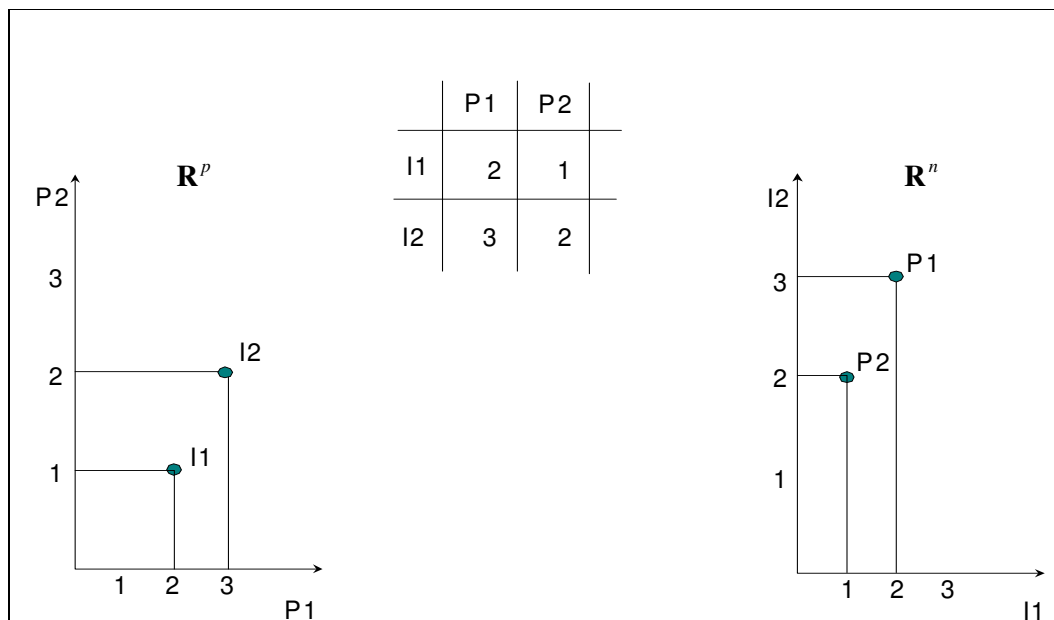


Fig. 2.1 - Representação das nuvens em  $R^p$  e  $R^n$ .

Quando o número de indivíduos e propriedades é elevado, não é possível detectar a estrutura da nuvem em estudo sem procurar diminuir a dimensão do espaço, objectivo comum dos vários métodos de Análise Factorial de Dados.

Esta redução de dimensão deve assegurar uma perda de informação mínima, isto é, o sub-espaço a pesquisar tem de garantir a menor deformação possível da nuvem inicial.

Seja uma nuvem constituída por  $n$  pontos  $x_i$  com coordenadas  $x_{ij}$  e massa  $m_i$ . O centro de gravidade da nuvem é dado por

$$g = \sum_{i=1}^n m_i x_i$$

A média ponderada pelas massas dos  $n$  indivíduos da variável  $j$ , dada por

$$g_j = \sum_{i=1}^n m_i x_{ij}$$

constitui a  $j$ -ésima componente do centro de gravidade da nuvem dos indivíduos. Em particular, se  $m_i = \frac{1}{n}$  vem

$$g_j = \frac{1}{n} \sum_{i=1}^n x_{ij} = \bar{x}_j$$

A inércia da nuvem em relação a um ponto  $a$  é dada por:

$$I_a = \sum_{i=1}^n m_i (x_i - a)^2$$

Note-se a equivalência entre a variância da variável  $j$  e a inércia em relação ao seu centro de gravidade

$$I_{g_j} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - g_j)^2$$

A matriz de inércia, dada por:

$$V = \sum_{i=1}^n m_i (x_i - g)^T (x_i - g)$$

é o equivalente geométrico da matriz variância-covariância (fazendo  $m_i = \frac{1}{n}$ ). Note-se que a inércia em relação ao centro de gravidade se pode escrever:

$$I_g = \text{tr}(V) = \sum_{i=1}^n v_{jj}$$

Seja  $W$  um sub-espço vectorial de  $R^p$  e  $W^\perp$  o seu suplementar ortogonal. Então qualquer vector  $x_i$  de  $R^p$  pode decompor-se, de forma única, na soma de 2 vectores:

$$x_i = y_i + z_i \quad , \quad x_i \in R^p; \quad y_i \in W; \quad z_i \in W^\perp$$

com  $y_i$  e  $z_i$  ortogonais ( $y_i \cdot z_i = 0$ ).

Se o centro de gravidade da nuvem coincidir com a origem temos

$$I_w = \sum_{i=1}^n m_i z_i^2$$

$$I_{w^\perp} = \sum_{i=1}^n m_i y_i^2$$

Como os dois espaços são ortogonais vem (Teorema de Pitágoras):

$$I_g = I_w + I_{w^\perp}$$

## ANÁLISE GERAL

### Análise em $R^p$

Pretende-se ajustar a nuvem de  $n$  pontos por um sub-espço vectorial munido da distância euclideana usual.

Comecemos por procurar a recta  $F_1$ , passando pela origem, que melhor ajusta a nuvem, segundo o critério dos mínimos quadrados. Seja  $u_1$ , um vector unitário sobre a recta, isto é, tal que

$$u_1^T u_1 = 1 \quad \text{ou} \quad \sum_{j=1}^p u_{1j}^2 = 1$$

Cada linha de  $X$  representa um ponto em  $R^p$ . A matriz  $Xu_1$  representa as projecções dos  $n$  pontos em  $F_1$  (vd. Fig. 2.2).

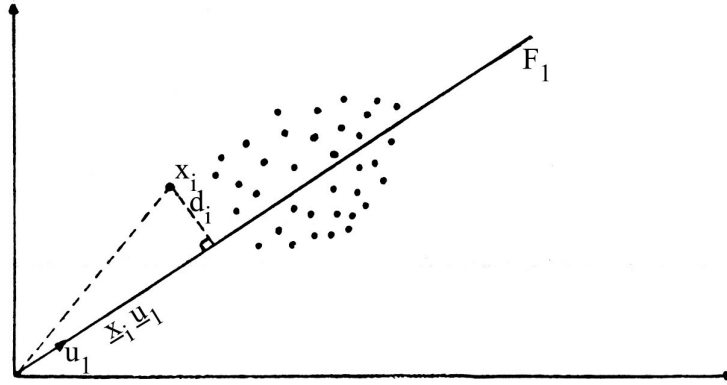


Fig. 2.2 - Projecção de um ponto num vector.

O quadrado da distância de um ponto  $x_i$  à origem decompõe-se na soma do quadrado da projecção em  $F_1$  ( $x_i u_1$ ) com o quadrado da distância a  $F_1$  ( $d_i$ ). Minimizar a perda de inércia é equivalente a minimizar a soma dos quadrados das distâncias a  $F_1$ , ou maximizar a soma dos quadrados das projecções em  $F_1$ , dada por

$$(X u_1)^T X u_1 = u_1^T X^T X u_1$$

Encontrar a recta que melhor se ajusta a nuvem traduz-se em maximizar a forma quadrática

$$u_1^T X^T X u_1 \text{ sujeita ao constrangimento } u_1^T u_1 = 1.$$

Para encontrar o máximo daquela forma quadrática formemos a lagrangeana  $L$  e derivemo-la em ordem a  $u_1$

$$L = u_1^T X^T X u_1 - \lambda_1 (u_1^T u_1 - 1)$$

$$\frac{dL}{du_1} = 2 X^T X u_1 - 2\lambda_1 u_1 = 0$$

$$X^T X u_1 - \lambda_1 u_1 = 0 \tag{2.1}$$

Conclui-se pois que  $u_1$  é um vector próprio de  $X^T X$  e  $\lambda_1$ , que lhe está associado, é o maior valor próprio. Este valor próprio é o máximo procurado como se pode ver pré-multiplicando (2.1) por  $u_1^T$

$$u_1^T X^T X u_1 - \lambda_1 u_1^T u_1 = 0$$

$$u_1^T X^T X u_1 = \lambda_1 \tag{2.2}$$

Procuramos agora o espaço a duas dimensões que melhor ajusta a nuvem. Trata-se de encontrar uma recta  $F_2$  definida pelo vector unitário  $u_2$  ( $u_2^T u_2 = 1$ ) perpendicular a  $u_1$  ( $u_2^T u_1 = 0$ ).

Neste caso a lagranjeana toma a forma

$$L = u_2^T X^T X u_2 - \lambda_2(u_2^T u_2 - 1) - \theta u_2^T u_1 \quad (2.3)$$

$$\frac{dL}{du_2} = 2 X^T X u_2 - 2\lambda_2 u_2 - \theta u_1 = 0 \quad (2.4)$$

Pré-multiplicando (2.4) por  $u_1^T$  vem:

$$\begin{aligned} 2 u_1^T X^T X u_2 - 2\lambda_2 u_1^T u_2 - \theta u_1^T u_1 &= 0 \\ 2 u_1^T X^T X u_2 - \theta &= 0 \end{aligned} \quad (2.5)$$

Pós-multiplicando (2.2) por  $u_1^T$  vem:

$$\begin{aligned} u_1^T X^T X u_1 u_1^T &= \lambda_1 u_1^T \\ u_1^T X^T X &= \lambda_1 u_1^T \end{aligned}$$

Substituindo em (2.5) vem:

$$2\lambda_1 u_1^T u_2 - \theta = 0$$

Como  $u_1^T u_2 = 0$ , da equação anterior resulta que o segundo parâmetro de Lagrange  $\theta$  é nulo. Em consequência (2.4) transforma-se em

$$\begin{aligned} 2 X^T X u_2 - 2\lambda_2 u_2 &= 0 \\ X^T X u_2 - \lambda_2 u_2 &= 0 \end{aligned}$$

Assim  $u_2$  é o vector próprio associado ao segundo maior valor próprio de  $X^T X$ . Generalizando mostra-se que

$$X^T X u_\alpha - \lambda_\alpha u_\alpha = 0$$

De notar que  $\text{tr}(X^T X) = \sum \lambda_\alpha$ .

Assim o sub-espaço vectorial de dimensão  $p_r < p$  que melhor ajusta a nuvem, segundo um critério de mínimos quadrados, é engendrado por uma base constituída pelos  $p_r$  vectores próprios associados aos  $p_r$  maiores valores próprios de  $X^T X$ .

### Análise em $R^n$ . Fórmulas de transição.

Cada coluna da matriz  $X$  representa agora um ponto em  $R^n$ , o que conduz a uma nuvem de  $p$  pontos.

Procuremos o vector unitário  $v_l$  que maximiza a soma dos quadrados das projecções dos pontos sobre a recta  $G_l$  a ele associado. Pretende-se agora maximizar a forma quadrática seguinte

$$(X^T v_l)^T (X^T v_l) = v_l^T X X^T v_l$$

De uma forma semelhante à utilizada para a análise em  $R^p$  verifica-se que  $v_l$  é o vector próprio correspondente ao máximo valor próprio de  $X X^T$ . No caso geral virá:

$$X X^T v_\alpha - \mu_\alpha v_\alpha = 0$$

Procuremos as relações entre os valores e vectores próprios de  $X^T X$  e  $X X^T$ . Da análise em  $R^p$  e  $R^n$ , resultou respectivamente:

$$X^T X u_\alpha - \lambda_\alpha u_\alpha = 0 \quad (2.6)$$

$$X X^T v_\alpha - \mu_\alpha v_\alpha = 0 \quad (2.7)$$

Pré-multiplicando a equação (2.6) por  $X$  vem

$$X X^T X u_\alpha - \lambda_\alpha X u_\alpha = 0 \quad (2.8)$$

A relação (2.8) mostra que  $X u_\alpha$  são vectores próprios de  $X X^T$ , sendo  $\lambda_\alpha$  os respectivos valores próprios. Comparando (2.6) e (2.7), verifica-se que  $\lambda_\alpha$  coincide com  $\mu_\alpha$ , porque são os valores próprios da mesma matriz.

Comparando (2.7) e (2.8) verifica-se que  $X u_\alpha$  e  $v_\alpha$  são proporcionais, pois são vectores próprios da mesma matriz:

$$v_\alpha = k_1 X u_\alpha \quad (2.9)$$

Pré-multiplicando (2.7) por  $X^T$  e comparando os resultados com (2.6) verifica-se analogamente a proporcionalidade entre  $u_\alpha$  e  $X^T v_\alpha$ :

$$u_\alpha = k_2 X^T v_\alpha \quad (2.10)$$

Pré-multiplicando (2.9) e (2.10) respectivamente por  $v_\alpha^T$  e  $u_\alpha^T$  temos

$$1 = k_1 v_\alpha^T X u_\alpha \Rightarrow 1 = k_1 a \quad (2.11)$$

$$1 = k_2 u_\alpha^T X^T v_\alpha \Rightarrow 1 = k_2 a$$

com

$$a = (v_\alpha X u_\alpha)^T = u_\alpha^T X^T v_\alpha$$

Donde resulta que  $k_1 = k_2 = k$ .

Pré-multiplicando (2.10) por  $v_\alpha^T X$ , vem

$$v_\alpha^T X u_\alpha = k v_\alpha^T X X^T v_\alpha$$

que, atendendo a (2.2), é equivalente a

$$v_\alpha^T X u_\alpha = k \lambda_\alpha$$

Substituindo em (2.11) vem

$$\frac{1}{k} = k \lambda_\alpha$$

$$\frac{1}{k^2} = \lambda_\alpha \quad \Rightarrow \quad k = \frac{1}{\sqrt{\lambda_\alpha}}$$

Donde, substituindo em (2.9) e (2.10), vem:

$$\boxed{\begin{aligned} v_\alpha &= \frac{1}{\sqrt{\lambda_\alpha}} X u_\alpha \\ u_\alpha &= \frac{1}{\sqrt{\lambda_\alpha}} X^T v_\alpha \end{aligned}} \quad (2.12)$$

Estas fórmulas de transição de um espaço a outro são de grande importância pois simplificam o estudo das duas nuvens. É suficiente a diagonalização de apenas uma das matrizes  $X^T X$  ou  $X X^T$ , normalmente a de menor dimensão.