Processamento digital de imagens

Agostinho Brito

Departamento de Engenharia da Computação e Automação
Universidade Federal do Rio Grande do Norte

28 de abril de 2016

Análise de componentes principais

- A análise de componentes principais, ou transformada de Karhunen-Loève, ou transformada de Hotelling, realiza uma transformação ortogonal em um conjunto de variáveis correlacionadas (interdependentes) em um outro conjunto de variáveis linearmente descorrelacionadas.
- Dá-se o nome de "componentes principais" porque no espaço transformado, as primeiras componentes concentram a maior parte da informação presente nos dados originais.

Formulação

- Seja $\bar{x} = [\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array}]^T$ um vetor representante de uma população de dados com n-dimensional.
- Define-se o vetor de média e a matriz de covariância dessa população como

$$\bar{m}_x = Ex \qquad C_x = E(x - \bar{m}_x)(x - \bar{m}_x)^T$$

- Para uma amostra de M elementos, a média e covariância da população podem ser aproximadas por

$$\bar{m}_x = \frac{1}{M}\sum_{k=1}^{M} \qquad C_x = \frac{1}{M}\sum_{k=1}^{M}(\bar{x}_k\bar{x}_k^T) - \bar{m}_k\bar{m}_k^T$$

- Ex: $\bar{x}_1 = (0,0,0)^T$, $\bar{x}_2 = (1,0,0)^T$, $\bar{x}_2 = (1,1,0)^T$ e $\bar{x}_4 = (1,0,1)^T$.

$$\bar{m}_x = [\begin{array}{ccc} 3 & 1 & 1 \end{array}]^T \quad C_x = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

Formulação (cont.)

- A matriz C_x é real e simétrica. Logo, possui autovalores reais e distintos $\lambda_i, i = 1, 2, \cdots, n$, com seus respectivos autovetores $\bar{e}_i$ associados.
- Assuma que estes autovalores são arranjados na ordem decrescente de valor, tais que $\lambda_j \geq \lambda_{j+1}$ para $j = 1, 2, \cdots, n-1$.
- Construa uma matriz A, onde cada linha contenha autovetores unitários associados com os autovalores da matriz C_x. A primeira linha de C_x deve conter o autovetor associado com o maior autovalor, restando na última linha o autovetor associado com o menor autovalor.
- A transformação

  $$\bar{y} = A(\bar{x} - \bar{m})$$

  é denominada de transformada de Hotelling. A população de vetores transformados possui média igual a zero. Sua matriz de covariância é diagonal, preenchida pelos autovalores da matriz C_x na ordem decrescente de seus valores.