



B3. PCA

Bổ sung thêm cho bài giảng

2019

Nội dung bổ sung



1. Ma trận hiệp phương sai
2. Principal Component Analysis

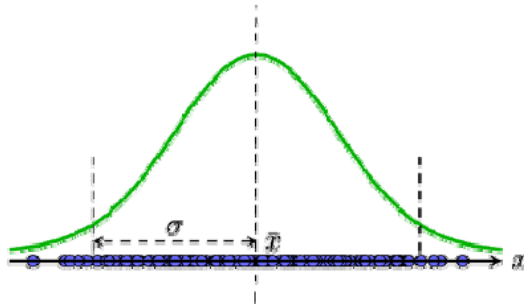


1. Ma trận hiệp phương sai

□ Kỳ vọng (*expectation*)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

trung bình cộng (*mean*)



□ Phương sai (*variance*) và độ lệch chuẩn (*standard deviation*)

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{trung bình khoảng cách đến kỳ vọng}$$

- σ : độ lệch chuẩn
- phương sai càng NHỎ thì các điểm dữ liệu càng gần kỳ vọng
- phương sai càng LỚN thì các điểm dữ liệu càng phân tán

1. Ma trận hiệp phương sai (tt.)



□ Vector cột $x_1, x_2, \dots, x_n \in \mathbb{R}^m$, ma trận $X = (x_1 \ x_2 \ \dots \ x_n) \in \mathbb{R}_{m,n}$

- ma trận trung tâm (*center matrix* \neq *centering matrix*) $\hat{X} \in \mathbb{R}_{m,n}$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{x}_{ij} = (x_{ij} - \bar{x})$$

$$\hat{X} = (X - \bar{x}) = ((x_1 - \bar{x}) \ (x_2 - \bar{x}) \ \dots \ (x_n - \bar{x}))$$

có thể tính trung bình trên mỗi cột

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \rightarrow \hat{x}_{ij} = (x_{ij} - \bar{x}_j)$$

- ma trận hiệp phương sai (*covariance matrix*) của X

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^T \cdot (x_i - \bar{x}) = \frac{1}{n} \hat{X}^T \cdot \hat{X} \equiv V$$



1. Ma trận hiệp phương sai (tt.)

□ Một số tính chất của ma trận hiệp phương sai V

- ma trận đối xứng
- ma trận nửa xác định dương
- hệ số không âm trên đường chéo: phương sai trên từng chiều
- hiệp phương sai v_{ij} ($i \neq j$): mối tương quan giữa x_i và x_j
- nếu V là ma trận đường chéo \Rightarrow hoàn toàn không tương quan



Nội dung bổ sung

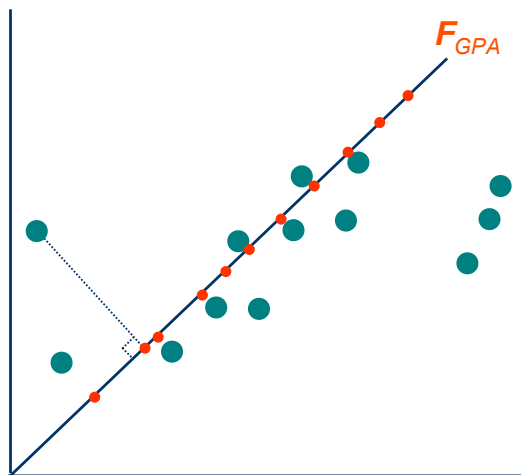


1. Ma trận hiệp phương sai
2. Principal Component Analysis

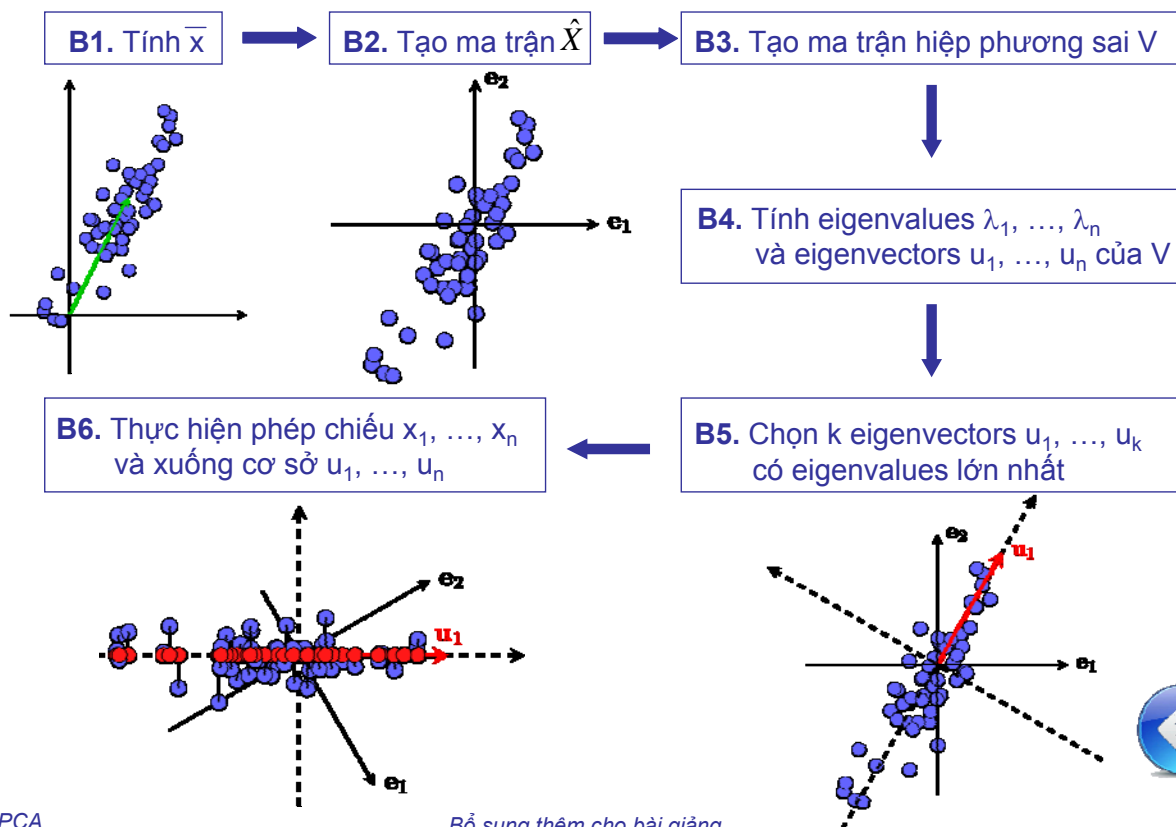
2. Principal Component Analysis (tt.)



- Tìm không gian đặc trưng mới F' tạo phân hoạch trên items tốt hơn không gian đặc trưng ban đầu F



2. Principal Component Analysis



2. Principal Component Analysis (tt.)



□ Hệ cơ sở – Tọa độ trong không gian vectơ V

Cơ sở “có thứ tự” B gồm các vectơ độc lập tuyến tính:

$$B = \{ u_1, u_2, \dots, u_n \}$$

$$\forall v \in V: \quad v = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_n u_n \quad \alpha_i \in \mathbb{R}$$

Mã trận cơ sở của không gian V: $B = \begin{pmatrix} u_1^T & u_2^T & \dots & u_n^T \end{pmatrix}$

Tọa độ của v theo B:

$$[v]_B = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

2. Principal Component Analysis (tt.)



□ Hệ cơ sở – Tọa độ trong không gian vectơ V

Cơ sở “có thứ tự” B' gồm các vectơ độc lập tuyến tính:

$$B = \{ u'_1, u'_2, \dots, u'_n \}$$

Mã trận chuyển đổi cơ sở từ B sang B':

$$(B \rightarrow B') = \begin{pmatrix} [u'_1]_B & [u'_2]_B & \dots & [u'_n]_B \end{pmatrix}$$

$(B \rightarrow B')$ khả nghịch

Công thức chuyển đổi tọa độ:

$$[v]_{B'} = (B \rightarrow B')^{-1} [v]_B$$

$$[v]_B = (B \rightarrow B') [v]_{B'}$$