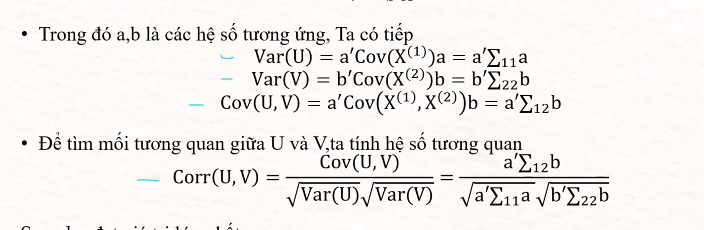
Tuần 5

**Phân tích tương quan chính tắc (CCA)**

1. **Chứng minh 4 công thức bên dưới: và**

****

1. **Phân biệt Sample Mean và Mean Vector**

| sample mean | kì vọng E, sigma |
| --- | --- |
| mẫu | quần thể |
| Phải lấy đủ mẫu mới có thể tính được vì không có hàm phân bố mẫu. | Không cần phải đi lấy đủ mẫu để tính vì đã có hàm phân bố xác suất/ phân bố mẫu (kì vọng ), hàm này cho chúng ta biết mẫu được phân bố như thế nào |

1. **Trình bày ý nghĩa hình học của Corr(U,V)**

Góc giữa 2 vector u,v bé nhất => cos(u,v) lớn nhất.

1. **Lập bảng phân biệt PCA và CCA và linear regression**

|  | **PCA** | **CCA** | **Linear Regression** |
| --- | --- | --- | --- |
| **Giảm số chiều** | Có | Có |  |
| **Tương quan các biến** | TÌm được các thành phần không tương quan | Tương quan giữa tập biến sao cho lớn nhất |  |
| **ưu** |  |  |  |
| **khuyết** |  |  |  |

**Phân tích dữ kiện (FA)**

1. **PCA vs FA**

|  | pca | fa | hồi quy tt | cca |
| --- | --- | --- | --- | --- |
|  | xấp xỉ tuyến tính vector gốc thành tổ hợp vector riêng ứng với trị riêng max |  |  |  |
| sự phụ thuộc tuyến tính | có | giữa x và các vector có sự phụ thuộc. | Có ở quan hệ giữa các biến độc lập và các biến phụ thuộc | có |
| phạm vi ứng dụng | Bài toán có dữ liệu lớn (số chiều lớn) hoặc dư thừa dữ liệu ảnh (pixel); 2 lớp và có hơn 30 thuộc tính  => rút trích thành phần quan trọng để biểu diễn dữ liệu lại.  Tiêu chuẩn: tỷ số % giá trị riêng L max chia với tổng các L (phương sai max tập trung tại những giá trị riêng nào thì giữ lại vector riêng/không gian con đó) |  |  |  |

1. **Trong 4 phương pháp chúng ta đã sử dụng thì vai trò của sample và population ở đâu ?**

|  | **LR** | **PCA** | **CCA** | **FA** |
| --- | --- | --- | --- | --- |
| dữ liệu | mẫu ở cực tiểu sai số | quần thể ở cực đại sigma variance, sigma covariance | quần thể ở cực đại sigma correlation | quần thể ở cực tiểu sai số |

**(sigma**: hàm phân bố xác suất**)**