KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax:(08) 38.350.096

ĐỒ ÁN MÔN HỌC

MÔN: PHƯƠNG PHÁP THỐNG KÊ DỮ LIỆU NHIỀU BIẾN

Chủ đề: Phân tích tương quan chính tắc

Nội dung

A. Lý thuyết:

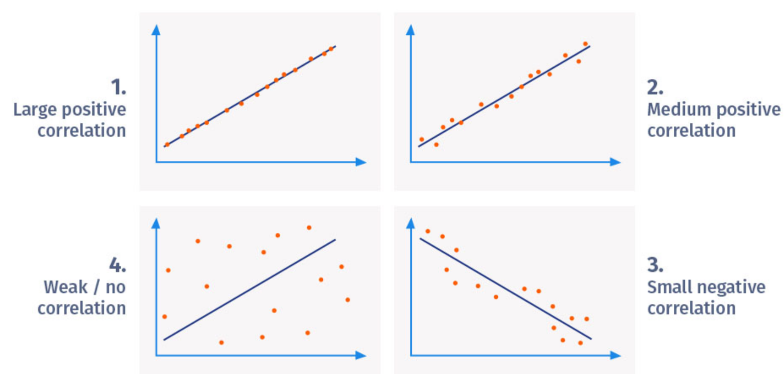
1. Mục đích:

- Với bài toán CCA thực chất ta muốn xác định cách thể hiện của 2 tập dữ liệu sao cho độ tương quan (correlation) là lớn nhất.
- Đầu tiên khi xem xét với 2 tập dữ liệu X:(x) ,Y:(y) 1 biến vd:(tiền lương, mức sống) ta sẽ xem thử liệu 2 tập dữ liệu này có mối quan hệ nào không dựa vào hệ số tương quan thông qua công thức.

$$Corr(X, Y) = Corr(x, y) = \frac{Cov(x,y)}{\sqrt{Var(x)} * \sqrt{Var(y)}}$$

Với $Cov(x, y) = E[(x - E(x))(x - E(y))]$

$$Var(x) = E[(x - E(x))^2] = Cov(x, x), \text{ tương tự với } y$$



- Hệ số tương quan (correlation coefficient) dao động từ -1 đến 1:
  + Khi $\approx 1 \; hay \approx -1$ ta nói 2 biến có hệ số tương quan lớn, sẽ có xu hướng phụ thuộc tuyến tính vào nhau theo hệ số dương, nếu x tăng thì y cũng tăng.
  + Khi $\approx$ 0 thì lúc này cả 2 biến đều xem như độc lập hay có mối tương quan rất thấp không thể từ biến này suy ra biến còn lại.

- Ở trên khi xét trường hợp 1 biến cho mỗi tập dữ liệu ta thấy việc xác định mối tương quan của 2 tập dữ liệu rất dễ dàng nhưng nếu là 2 tập dữ liệu đa biến thì mọi việc lại không đơn giản như vậy
- Xét 2 vector ngẫu nhiên $X = (x_1, x_2, x_3, ...)^T$ và $Y = (y_1, y_2, y_3, ...)^T$
- Hệ số tương quan lúc này là 1 ma trận đối xứng

  $$Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)} * \sqrt{Var(Y)}}$$

  $corr_{ij}$ thể hiện mối tương quan giữa biến x_i và y_j và chỉ xem xét giữa các biến ngẫu nhiên trong 2 vector ngẫu nhiên, dựa vào ma trận Corr(X, Y) ta khó có thể suy ra được mối quan hệ của 2 tập biến ngẫu nhiên hay có thể nói là không thể.
- Từ đó nhu cầu về xem xét mối quan hệ giữa 2 tập biến được đặt ra và phân tích tương quan chính tắc - (Canonical Correlation Analysis) được ra đời.
- Liên hệ từ bài toán Hồi quy tuyến tính (Linear Regression), ta tìm một tổ hợp tuyến tính của X là $a^T X$ với $a = (a_1, a_2, a_3, ...)^T$ để có thể mô tả vector ngẫu nhiên và giúp cho việc dự đoán, tương tự với Y là $b^T Y$ với $b = (b_1, b_2, b_3, ...)^T$ từ 2 vector ngẫu nhiên ta đã đưa về bài toán ban đầu là phân tích hệ số tương quan của 2 biến, giúp bài toán dễ dàng tiếp cận hơn.
- Và lúc này mỗi vector ngẫu nhiên sẽ được đại diện bởi 1 vector tổ hợp tuyến tính và tính toán sao cho hệ số tương quan là lớn nhất có thể, để thể hiện được tương quan chính tắc của 2 vector ngẫu nhiên.
- Tóm lại phân tích tương quan chính tắc chính là phân

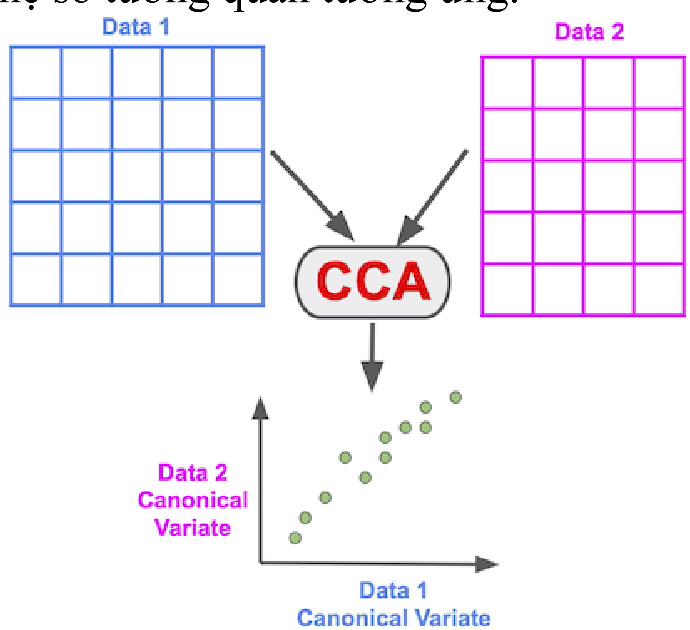
tích sự tương quan của 2 tập biến ngẫu nhiên (hay 2 tập dữ liệu) và chọn các vector tổ hợp tuyến tính đi kèm một số điều kiện và làm cho hệ số tương quan là lớn nhất có thể.

2. Phát biểu bài toán

- Cho 2 tập dữ liệu được biểu diễn bằng 2 ma trận $(p\ x\ n)X^{(1)} = (X_1^{(1)}, X_2^{(1)}, X_3^{(1)}, ..., X_p^{(1)})^T$

và $(q\ x\ n)X^{(2)} = (X_1^{(2)}, X_2^{(2)}, X_3^{(2)}, ..., X_q^{(2)})^T$

- Với $X_i^{(1)} = (x_{i1}^{(1)}, x_{i2}^{(1)}, x_{i3}^{(1)}, ..., x_{in}^{(1)}),\ i = 1...p$
- $X_i^{(2)} = (x_{i1}^{(2)}, x_{i2}^{(2)}, x_{i3}^{(2)}, ..., x_{in}^{(2)}),\ i = 1...q$
- Kết quả là cặp vector $a = (a_1, a_2, a_3, ... a_p)^T$, $b = (b_1, b_2, b_3, ... b_q)^T$ làm cho hệ số tương quan giữa 2 tập dữ liệu là lớn nhất và hệ số tương quan tương ứng.



KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax:(08) 38.350.096

3. Phương pháp

- Xét 2 vector ngẫu nhiên:

$$(p\ x\ 1)X^{(1)} = (X^{(1)}{}_1, X^{(1)}{}_2, X^{(1)}{}_3, ..., X^{(1)}{}_p)^T$$

$$(q\ x\ 1)X^{(2)} = (X^{(2)}{}_1, X^{(2)}{}_2, X^{(2)}{}_3, ..., X^{(2)}{}_q)^T$$

- Các đại lượng sẽ được ký hiệu:

$$E\left(X^{(1)}\right) = \mu^{(1)}(p\ x\ 1),\ E\left(X^{(2)}\right) = \mu^{(2)}(q\ x\ 1)$$

$$Cov(X^{(1)}) = \Sigma_{11}(p\ x\ p),\ Cov(X^{(2)}) = \Sigma_{22}(q\ x\ q)$$

$$Cov(X^{(1)}, X^{(2)}) = \Sigma_{12} = \Sigma_{21}{}^T\ (px\ q)$$

- Ma trận Σ_{12} thể hiện sự tương quan của các biến trong 2 vector ngẫu nhiên (2 tập dữ liệu). Khi q p lớn việc phân tích sự tương quan của 2 tập dữ liệu là bất khả thi, vì thế phân tích tương quan chính tắc ra đời để xem xét sự tương quan của 2 tập dữ liệu nhiều biến bằng 1 cách tiếp cận khác thay vì dựa vào Σ_{12}.
- Bằng cách sử dụng tổ hợp tuyến tính của vector ngẫu nhiên ta đơn giản bài toán thành việc tính toán hệ số tương quan (correlation coefficient) giữa 2 biến.
- Ta đặt: $U = a^T X^{(1)},\ V = b^T X^{(2)}$

  với $a = (a_1, a_2, a_3, ...\ a_p)^T,\ b = (b_1, b_2, b_3, ...\ b_q)^T$
- Dựa vào tính chất tính toán các đại lượng của 1 tổ hợp tuyến tính đối với vector ngẫu nhiên ta được:

$$E(U) = a^T E(X^{(1)}) = a^T \mu^{(1)} Var(U) = a^T Cov(X^{(1)}) a = a^T \Sigma_{11} a$$

$$E(V) = b^T E(X^{(2)}) = b^T \mu^{(2)}$$
$$Var(V) = b^T Cov(X^{(2)}) b = b^T \Sigma_{22} b$$
$$Cov(U, V) = a^T Cov(X^{(1)}, X^{(2)}) b = a^T \Sigma_{12} b$$

- Hệ số tương quan của U và V lúc này là:

$$Corr(U, V) = \frac{Cov(U,V)}{\sqrt{Var(U)} * \sqrt{Var(V)}} = \frac{a^T \Sigma_{12} b}{\sqrt{a^T \Sigma_{11} a} * \sqrt{b^T \Sigma_{22} b}}$$

nhận thấy Corr(U, V) phụ thuộc vào a và b vì thế ta sẽ tìm a, b sao cho hệ số tương quan của U và V là lớn nhất, từ đó sẽ xác định được sự tương quan giữa 2 tập biến đang xét.

- Lưu ý rằng vì U và V là đang xét trên phương diện 1 biến n

$$Corr(U, V) = \frac{\frac{1}{n} d_U . \frac{1}{n} d_V}{\sqrt{\frac{1}{n^2} * (d_U^T d_U)} \sqrt{\frac{1}{n^2} * (d_V^T d_V)}} = \frac{d_U . d_V}{L_{d_U} * L_{d_V}} = cos(\theta)$$

- Trong đó d_U, d_V lần lượt là deviation vector của U và V, L_{d_U}, L_{d_V} là độ dài vector d_U, d_V. Vì thế hệ số tương quang sẽ không thay đổi với bất kỳ tỉ lệ scale của U và V.

Giải bài toán:

- Ta có $Corr(U, V) = \frac{a^T \Sigma_{12} b}{\sqrt{a^T \Sigma_{11} a} * \sqrt{b^T \Sigma_{22} b}}$

- Đặt $a = \Sigma_{11}^{-\frac{1}{2}} * c \Rightarrow c = \Sigma_{11}^{\frac{1}{2}} a$

  $b = \Sigma_{22}^{-\frac{1}{2}} * d \Rightarrow d = \Sigma_{22}^{\frac{1}{2}} b$

$$\Rightarrow Corr(U, V) = \frac{c^T \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} d}{\sqrt{c^T c} * \sqrt{d^T d}} = \frac{c^T M d}{\sqrt{c^T c} * \sqrt{d^T d}} = \rho$$

$$Với\ M = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$$

- Áp dụng bất đẳng thức Cauchy-Schwarz cho 2 vector c và Md ta được: $(c^T c)(d^T M^T M d) \geq (c^T M d)^2$
  $dấu\ =\ xảy\ ra\ khi\ c = kMd,\ k$ là $hằng\ số$ hay
  $$\rho^2 = \frac{(c^T M d)^2}{(c^T c)\ (d^T d)} \leq \frac{(c^T c)(d^T M^T M d)}{(c^T c)\ (d^T d)} = \frac{(d^T M^T M d)}{(d^T d)}$$
- Lúc này bài toán đưa về tìm max của $\frac{(d^T M^T M d)}{(d^T d)}$ là một bài toán rất quen thuộc trong phân tích thành phần chính - (PCA)

- Chứng minh bài toán PCA: tìm max của $\frac{(d^T M^T M d)}{(d^T d)}$

  Vì $M^T M$(k x k) là một ma trận đối xứng, áp dụng phân rã ma trận đối xứng $M^T M = P \Lambda P^T,\ PP^T = I$

  $P = \left(p_1, p_2, p_3, \ldots, p_k\right)^T$ gồm các vector riêng ứng với giá trị riêng được.

  Λ là ma trận chéo chứa các giá trị riêng từ lớn đến bé. ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$).

$$\frac{(d^T M^T M d)}{(d^T d)} = \frac{d^T P \Lambda P^T d}{d^T d} = \frac{x^T \Lambda x}{x^T x}\ với\ x = P^T d\ (k\ x\ 1)$$

Vì Λ là ma trận đường chéo chứa các giá trị riêng của $M^T M$ ta được

$$\frac{x^T\Lambda x}{x^T x} = \frac{\sum_{i=1}^{q}\lambda_i x_i^2}{\sum_{i=1}^{q} x_i^2} \leq \lambda_1 \frac{\sum_{i=1}^{q} x_i^2}{\sum_{i=1}^{q} x_i^2} = \lambda_1$$

Vậy để cho $\frac{x^T\Lambda x}{x^T x}$ đạt giá trị lớn nhất khi $d = p_1$

$x = P^T d = P^T p1 = [1, 0, 0, 0..., 0]^T$

Suy ra $\frac{(d^T M^T M d)}{(d^T d)} = \frac{x^T\Lambda x}{x^T x} = \lambda_1$ khi $d = p_1$ điều phải chứng minh.

- Quay về lại bài toán CCA: ta có max của $\frac{(d^T M^T M d)}{(d^T d)} = \lambda_1$ là giá trị riêng lớn nhất của ma trận $M^T M = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$, đồng thời $d = e_1$ là vector riêng của ma trận $M^T M$ ứng với giá trị riêng λ_1.

$$\Rightarrow \rho^2 \leq \lambda_1 \Rightarrow \rho \leq \sqrt{\lambda_1}$$

- Để đạt trường hợp lớn nhất thì dấu bằng phải xảy ra.
- Dấu bằng xảy ra khi $c = kMd$

$\Leftrightarrow M^T c = kM^T M d$, *ta lại có* $d = e_1 \Rightarrow M^T M d = \lambda_1 d$

$\Rightarrow M^T c = k\lambda_1 d \Leftrightarrow MM^T c = \lambda_1 kMd \Leftrightarrow MM^T c = \lambda_1 c$

Vậy $c = f_1$ chính là vector riêng của ma trận ứng giá trị riêng λ_1

Lúc này:

$$Max(Corr(U,\ V))\ =\frac{a^T\Sigma_{12}b}{\sqrt{a^T\Sigma_{11}a}*\sqrt{b^T\Sigma_{22}b}}=\sqrt{\lambda_1}$$

- Tổng kết lại, để hệ số tương quan lớn nhất thì: $a\ =\Sigma_{11}^{-\frac{1}{2}}*c,\ b\ =\Sigma_{22}^{-\frac{1}{2}}*d$
  Với $c\ =f_1, d\ =e_1$ ứng với giá trị riêng λ_1 lớn nhất của ma trận MM^T và M^TM, ta gọi a b là cặp biến chính tắc mô tả tổ hợp tuyến tính của $X^{(1)}$ và $X^{(2)}$ là U và V và gọi $Corr(U,\ V)$ là độ tương quan chính tắc.

- Ta có các định nghĩa sau về phân tích tương quan chính tắc:
  + Cặp biến chính tắc thứ nhất là một cặp tổ hợp tuyến tính $U_1,\ V_1$, với phương sai đơn vị sao cho cực đại hóa mối tương quan $Corr(U_1,\ V_1)\ =\lambda_1$
  + Cặp biến chính tắc thứ hai cũng là một cặp tổ hợp tuyến tính $U_2,\ V_2$ với phương sai đơn vị và cực đại hóa mối tương quan $Corr(U_2,\ V_2)\ =\lambda_2$ kèm theo điều kiện là không có bất kỳ mối tương quan nào giữa cặp biến chính tắc thứ nhất.
  + Cặp biến thứ k là cặp biến thứ k $U_k,\ V_k$ với phương sai đơn vị và cực đại hóa mối tương quan $Corr(U_k,\ V_k)\ =\lambda_k$ với điều kiện là không có bất kỳ mối tương quan nào giữa k - 1 cặp

biến chính tắc.

- Các điều kiện được thể hiện qua phương trình sau:

$$Var(U_k) = Var(V_k) = 1$$

$$Cov(U_k, U_l) = Corr(U_k, U_l) = 0$$

$$Cov(V_k, V_l) = Corr(V_k, V_l) = 0$$

$$Cov(U_k, V_l) = Corr(U_k, V_l) = 0$$

$$Với\, k \neq l$$

- Chứng minh:

$$Cov(U_k, U_l) = Cov(a_k^T X^{(1)}, a_l^T X^{(1)})$$

$$= a_k^T Cov(X^{(1)}, X^{(1)}) a_l = a_k^T \Sigma_{11} a_l = f_k^T \Sigma_{11}^{\frac{-1}{2}} \Sigma_{11} \Sigma_{11}^{\frac{-1}{2}} f_l$$

$$= f_k^T f_l = 0 \text{ với } k \neq l, \text{ tương tự với } Cov(V_k, V_l)$$

$$Cov(U_k, V_l) = Cov(a_k^T X^{(1)}, b_l^T X^{(2)})$$

$$= a_k^T Cov(X^{(1)}, X^{(2)}) b_l = a_k^T \Sigma_{12} b_l = f_k^T \Sigma_{11}^{\frac{-1}{2}} \Sigma_{12} \Sigma_{22}^{\frac{-1}{2}} e_l$$

$$= f_k^T M e_l = \left(kMe_k\right)^T M e_l = k e_k^T \left(M^T M e_l\right) = k\lambda_l e_k^T e_l = 0$$

<u>(*) Vấn đề chuẩn hóa:</u>

- Xét $M^T M = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ chứng minh với MM^T tương tự

$$\left| \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} - \lambda I \right| = 0$$

$$\Leftrightarrow \left|\Sigma_{22}^{\frac{1}{2}}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{22}^{-\frac{1}{2}} - \lambda I\right| = 0$$

$$\Leftrightarrow \left|\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} - \lambda I\right| = |K - \lambda I| = 0$$

Vậy 2 ma trận có cùng trị riêng.

- Để thuận tiện ta xét $X^{(1)}, X^{(2)}$ đã được centralized, khi đó chuẩn hóa sẽ tương đương
- $Z^{(1)} = V^{(1)^{-\frac{1}{2}}}X^{(1)},\ Z^{(2)} = V^{(2)^{-\frac{1}{2}}}X^{(2)}\ với\ V = diag(\Sigma)$
- Khi đó ma trận K cho $Z^{(1)}, Z^{(2)}$ sẽ là:

$$K = \left(V^{(2)^{-\frac{1}{2}}}\Sigma_{21}V^{(1)^{-\frac{1}{2}}}\right)\left(V^{(1)^{\frac{1}{2}}}\Sigma_{11}^{-1}V^{(1)^{\frac{1}{2}}}\right)$$
$$\left(V^{(1)^{-\frac{1}{2}}}\Sigma_{12}V^{(2)^{-\frac{1}{2}}}\right)\left(V^{(2)^{\frac{1}{2}}}\Sigma_{22}^{-1}V^{(2)^{\frac{1}{2}}}\right)$$
$$= V^{(2)^{-\frac{1}{2}}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}V^{(2)^{\frac{1}{2}}}$$

$$áp\ dụng\ cm\ tương\ tự$$

$$\left|V^{(2)^{-\frac{1}{2}}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}V^{(2)^{\frac{1}{2}}} - \lambda I\right| = 0$$

$\Rightarrow \left|\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} - \lambda I\right| = 0$ vậy khi đó có thể kết luận, chuẩn hóa không làm thay đổi giá trị hệ số tương quan, cũng chính là giá trị riêng.

<u>Cặp vector chính tắc chuẩn hóa</u>:

- $U = a^TX^{(1)} = c^T\Sigma_{11}^{-\frac{1}{2}}X^{(1)} = c^T\Sigma_{11}^{-\frac{1}{2}}V^{(1)^{\frac{1}{2}}}Z^{(1)} = a_zZ^{(1)}$

- $V = b^T X^{(2)} = d^T \Sigma_{22}^{-\frac{1}{2}} X^{(2)} = d^T \Sigma_{22}^{-\frac{1}{2}} V^{(2)\frac{1}{2}} Z^{(2)} = b_z Z^{(2)}$
- Nhưng để đánh giá đúng về đóng góp của các biến việc chuẩn hóa là cần thiết để có cái nhìn thống kê về dữ liệu.

(*) Đánh giá các cặp vector chính tắc chuẩn hóa:

- xét $Z^{(1)}(p\ x\ n),\ Z^{(2)}(q\ x\ n)$ ta có:

$Z^{(1)} = A^{-1}U,\ Z^{(2)} = B^{-1}U\ ,\ A(p\ x\ p), B(q\ x\ q)$ chứa các vector chính tắc nhận làm vector cột

$$\Sigma_{11} = A^{-1}Cov(U,\ U)\left(A^{-1}\right)^T = A^{-1}\left(A^{-1}\right)^T = \sum_{i=1}^{p} a'_i(a'^T_i)$$

$$\Sigma_{22} = B^{-1}Cov(V,\ V)\left(B^{-1}\right)^T = B^{-1}\left(B^{-1}\right)^T = \sum_{i=1}^{q} b'_i(b'^T_i)$$

trong đó a'_i, b'_i là vector cột của ma trận A^{-1}, B^{-1}

$$\Sigma_{12} = A^{-1}Cov(U,\ V)\left(B^{-1}\right)^T = \sum_{i=1}^{p} \sqrt{\lambda_i}(a'_i)(b'^T_i)$$

- Khi đó để đánh giá k cặp vector chính tắc tốt như thế nào, người ta sẽ đánh giá dựa vào các vector chính tắc tái tạo lại các ma trận hiệp phương sai tốt như thế nào $\Sigma_{11}, \Sigma_{22}, \Sigma_{12}$.
- Tuy nhiên vì từ k + 1 đến p các hệ số $\sqrt{\lambda_i}$ sẽ gần tiến đến 0 và không tác động nhiều nên ta lưu ý vào Σ_{11}, Σ_{22}
- Ta có định nghĩa sau, tổng của các phương sai của

$Z^{(1)}$ được giải thích bởi k cặp vector chính tắc là:

$$R^2_{z^{(1)}}|U_1, U_2, U_3 \ldots U_k = \frac{tr\left(\sum_{i=1}^{k} a'_i (a'^T_i)\right)}{tr\left(\Sigma_{11}\right)}$$

$$= \frac{\sum_{i=0}^{k}\sum_{j=0}^{p} r^{(2)}_{U_i, z^{(1)}_j}}{p}$$

tương tự với $Z^{(2)}$

$$R^2_{z^{(2)}}|V_1, V_2, V_3 \ldots V_k = \frac{tr\left(\sum_{i=1}^{k} b'_i (b'^T_i)\right)}{tr\left(\Sigma_{22}\right)}$$

$$= \frac{\sum_{i=0}^{k}\sum_{j=0}^{q} r^{(2)}_{V_i, z^{(2)}_j}}{q}$$

với $r^{(2)}_{U_i, z^{(1)}_j}$, $r^{(2)}_{V_i, z^{(2)}_j}$ là hệ số tương quan bình phương.

<u>Chứng minh:</u>

$$Cov(Z^{(1)}, U) = Cov(A^{-1}U, U) = A^{-1}Cov(U, U) = A^{-1}$$

$$\mathbf{A}^{-1} = [\hat{\mathbf{a}}^{(1)}, \hat{\mathbf{a}}^{(2)}, \ldots, \hat{\mathbf{a}}^{(p)}] = \begin{bmatrix} r_{\hat{U}_1, z_1^{(1)}} & r_{\hat{U}_2, z_1^{(1)}} & \cdots & r_{\hat{U}_p, z_1^{(1)}} \\ r_{\hat{U}_1, z_2^{(1)}} & r_{\hat{U}_2, z_2^{(1)}} & \cdots & r_{\hat{U}_p, z_2^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{U}_1, z_p^{(1)}} & r_{\hat{U}_2, z_p^{(1)}} & \cdots & r_{\hat{U}_p, z_p^{(1)}} \end{bmatrix}$$

dễ nhận thấy $tr\left(a'_i(a'^T_i)\right) = \sum_{j=0}^{p} r^{(2)}_{U_i, z^{(1)}_j}$

tương tự với $Z^{(2)}$

- **<u>Giải pháp cho quần thể:</u>**

  Xét 2 vector ngẫu nhiên $X^{(1)}, X^{(2)}$như trên. Để thuận tiện ta xem xét vector X hợp bởi $X^{(1)}, X^{(2)}$:

$$\underbrace{X}_{((p+q)\,X\,1)} = \begin{bmatrix} X^{(1)} \\ \dots\dots. \\ X^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \dots\dots. \\ X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix}$$

(Giả sử p <= q)

Vector trung bình:

$$\underbrace{\mu}_{((p+q)X1)} = E(X) = \begin{bmatrix} E(X^{(1)}) \\ \dots\dots\dots.. \\ E(X^{(2)}) \end{bmatrix} = \begin{bmatrix} \mu^{(1)} \\ \dots\dots \\ \mu^{(2)} \end{bmatrix}$$

Ma trận hiệp phương sai:

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax:(08) 38.350.096

$$\Sigma_{(p+q)X(p+q)} = E(X-\mu)(X-\mu)^T$$

$$= \begin{bmatrix} E\left(X^{(1)}-\mu^{(1)}\right)\left(X^{(1)}-\mu^{(1)}\right)^T & \vdots & E\left(X^{(1)}-\mu^{(1)}\right)\left(X^{(2)}-\mu^{(2)}\right)^T \\ \cdots & \cdots & \cdots \\ E\left(X^{(2)}-\mu^{(2)}\right)\left(X^{(1)}-\mu^{(1)}\right)^T & \vdots & E\left(X^{(2)}-\mu^{(2)}\right)\left(X^{(2)}-\mu^{(2)}\right)^T \end{bmatrix}$$

$$= \begin{bmatrix} \underbrace{\Sigma_{11}}_{(pXp)} & \vdots & \underbrace{\Sigma_{12}}_{(pXq)} \\ \cdots & \cdots & \cdots \\ \underbrace{\Sigma_{21}}_{(qXp)} & \vdots & \underbrace{\Sigma_{22}}_{(qXq)} \end{bmatrix}$$

Với:

$$U = a^T X^{(1)}$$
$$V = b^T X^{(2)}$$

Là các tổ hợp tuyến tính tương ứng của $X^{(1)}, X^{(2)}$.

Ta được:

Cặp biến chính tắc thứ nhất:

$$a = \Sigma_{11}^{\frac{-1}{2}} f_1$$

$$b = \Sigma_{22}^{\frac{-1}{2}} e_1$$

$$\Rightarrow U_1 = f_1^T \Sigma_{11}^{\frac{-1}{2}} X^{(1)}, V_1 = e_1^T \Sigma_{22}^{\frac{-1}{2}} X^{(2)}$$

Với giá trị lớn nhất là

$$Corr(U_1, V_1) = \sqrt{\lambda_1}$$

…

Tương tự với cặp biến chính tắc thứ k, k=2,3,4,..,p:

$$U_k = f_k^T \Sigma_{11}^{\frac{-1}{2}} X^{(1)}, \quad V_k = e_k^T \Sigma_{22}^{\frac{-1}{2}} X^{(2)}$$

Giữa tất cả tổ hợp tuyến tính mà không có mối liên hệ(mối tương quan) với các cặp biến thứ 1,2, … , k-1 trước đó.
Có độ tương quan chính tắc:

$$Corr(U_k, V_k) = \sqrt{\lambda_k}$$

Ở đây:

+ $\lambda_1 > \lambda_2 > ... > \lambda_p$ là giá trị riêng lớn nhất là giá trị riêng lớn nhất của của ma trận đối xứng $M^T M = S_{22}^{-\frac{1}{2}} S_{21}\ S_{11}^{-1} S_{12}\ S_{22}^{-\frac{1}{2}}$ (qxq) (nếu q<p) là $MM^T = S_{11}^{-\frac{1}{2}} S_{12}\ S_{22}^{-1} S_{21}\ S_{11}^{-\frac{1}{2}}$ (pxp) nếu (p<q).
+ $e_1, e_2, ..., e_p$ là các vector riêng ứng với p giá trị riêng lớn nhất MM^T
+ $f_1, f_2, ..., f_p$ là các vector riêng ứng với p giá trị riêng lớn nhất của ma trận MM^T

- **<u>Giải pháp cho mẫu:</u>**
Xét mẫu ngẫu nhiên của n quan sát trên mỗi tập biến $X^{(1)}, X^{(2)}$. Để thuận tiện ta gộp thành ma trận dữ liệu X:

$$X = [X^{(1)} \ \vdots \ X^{(2)}]$$

$$= \left(\begin{matrix} x_{11}^{(1)} & x_{12}^{(1)} & x_{13}^{(1)} & \cdots & x_{1p}^{(1)} \\ & \vdots & & \ddots & \vdots \\ x_{n1}^{(1)} & x_{n2}^{(1)} & x_{n3}^{(1)} & \cdots & x_{np}^{(1)} \end{matrix} \middle| \begin{matrix} x_{11}^{(2)} & x_{12}^{(2)} & x_{12}^{(2)} & \cdots & x_{1q}^{(2)} \\ & \vdots & & \ddots & \vdots \\ x_{n1}^{(2)} & x_{n2}^{(2)} & x_{n3}^{(2)} & \cdots & x_{nq}^{(2)} \end{matrix}\right) = \begin{bmatrix} x_1^{(1)T} & \vdots & x_1^{(2)T} \\ \vdots & \vdots & \vdots \\ x_n^{(1)T} & \vdots & x_n^{(2)T} \end{bmatrix}$$

Vector trung bình mẫu:

$$\underbrace{\bar{x}}_{((p+q)X1)} = \begin{bmatrix} \bar{x}^{(1)} \\ \dots\dots\dots\dots \\ \bar{x}^{(2)} \end{bmatrix}$$

Ma trận hiệp phương sai mẫu:

$$\underset{((p+q)X(p+q))}{S} = \begin{bmatrix} \underbrace{S_{11}}_{(pXp)} & \vdots & \underbrace{S_{12}}_{(pXq)} \\ \dots\dots\dots & \dots & \dots\dots\dots \\ \underbrace{S_{21}}_{(qXp)} & \vdots & \underbrace{S_{22}}_{(qXq)} \end{bmatrix}$$

Với: $S_{kl} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j^{(k)} - x^{-(k)})(x_j^{(k)} - x^{-(k)})^T$

Với kết quả tương tự như tổng thể, ta có được:

$Corr(\hat{U}_k, \hat{V}_k) = \sqrt{\hat{\lambda}_k}$, với k=1,2,..,k-1

Với $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_p$ là giá trị riêng lớn nhất của của ma trận đối xứng $M^TM = S_{22}^{-\frac{1}{2}}S_{21}\ S_{11}^{-1}S_{12}\ S_{22}^{-\frac{1}{2}}$ (q x q) (Nếu q < p) là $MM^T = S_{11}^{-\frac{1}{2}}S_{12}\ S_{22}^{-1}S_{21}\ S_{11}^{-\frac{1}{2}}$ (p x p) nếu (p < q).

$\hat{\lambda}_1^2 > \hat{\lambda}_2^2 > ... > \hat{\lambda}_p^2$ cũng chính là các đại lượng tương quan mẫu.

Cùng cặp biến chính tắc thứ k tương ứng:

$$\hat{U}_k = \hat{f}_k^{\,T} \Sigma_{11}^{\frac{1}{2}} X^{(1)}, \hat{V}_k = \hat{e}_k^{\,T} \Sigma_{22}^{\frac{1}{2}} X^{(2)}$$

- **<u>Giải thuật:</u>**

Xét 2 vector ngẫu nhiên:

$$(p\ x\ 1)X^{(1)} = (X^{(1)}{}_1, X^{(1)}{}_2, X^{(1)}{}_3, ..., X^{(1)}{}_p)^T$$

$$(q\ x\ 1)X^{(2)} = (X^{(2)}{}_1, X^{(2)}{}_2, X^{(2)}{}_3, ..., X^{(2)}{}_q)^T$$

Min(p, q) = r giả sử min(p, q) = p

Bước 1:

- Chuẩn hóa các biến phân tích. Khi xử lý dữ liệu thô, đôi khi gặp nhiều dạng dữ liệu lạ khó phân tích. Để dễ tiếp cận ta đưa dữ liệu về một phân phối chuẩn

$$X^{(1)}{}_i = \frac{X^{(1)}{}_i - \underline{x}_i^{(1)}}{\sigma_i^{(1)}},\ X^{(2)}{}_i = \frac{X^{(2)}{}_i - \underline{x}_i^{(2)}}{\sigma_i^{(2)}}$$

- Với i=1,2,..,n .
  $\underline{x}_i^{(1)}, \underline{x}_i^{(2)}\ \sigma_i^{(1)}, \sigma_i^{(2)}$ lần lượt là kỳ vọng và phương sai của $X^{(1)}{}_i, X^{(2)}{}_i$
- Việc chuẩn hóa hay không sẽ không ảnh hưởng tới kết quả tương quan chính tắc.

Bước 2: Tính các ma trận (giả sử trường hợp r = p)

- $\Sigma_{11}^{-\frac{1}{2}}(p\ x\ p),\ \Sigma_{12} = \Sigma_{21}^T(p\ x\ q),\ \Sigma_{22}^{-1}(q\ x\ q)$

- $\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{22}^{-\frac{1}{2}}M^T\Sigma_{11}^{\frac{1}{2}}(q\ x\ p)$
- $\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} = MM^T(p\ x\ p)$

Bước 3:

- Tìm p giá trị riêng λ của ma trận MM^T qua phương trình

$$det\left(MM^T - \lambda I\right) = 0$$

Bước 4:

- Tìm $f_1, f_2, f_3, ..., f_p$ là các vector riêng ứng với p giá trị riêng của ma trận MM^T với $\left|f_i\right| = 1$
- $a_i = \Sigma_{11}^{-\frac{1}{2}} * f_i$
- $a = \Sigma_{11}^{-\frac{1}{2}} * c,\ b = \Sigma_{22}^{-\frac{1}{2}} * d$
- Áp dụng $MM^Tf_i = \lambda f_i \Rightarrow \left(M^TM\right)\left(M^Tf_i\right) = \lambda\left(M^Tf_i\right)$ vậy λ cũng là giá trị riêng của ma trận M^TM và nhận M^Tf_i là vector riêng.

  $e_i = M^Tf_i \Rightarrow b_i = \Sigma_{22}^{-\frac{1}{2}}e_i = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}f_i = \Sigma_{22}^{-1}\Sigma_{21}a_i$

  và scale sao cho $b_i^T\Sigma_{22}b_i = 1$ vì ở dạng toàn phương nên việc scale sẽ đơn giản khi làm việc với e_i, tiết kiệm thời gian tính toán
- Và mỗi cặp vector chính tắc a_i, b_i nhận $\sqrt{\lambda_i}$ là hệ số tương quan chính tắc.

4. **Ý nghĩa hình học:**

- Ý nghĩa hình học: đối với sample xét không gian n chiều, việc phân tích tương quan chính tắc chính là tìm tổ hợp của 2 vector $X_1^{(1)}$, $X_2^{(1)}$ nằm trong siêu phẳng được tạo bởi p vector $X_1^{(2)}$ và q vector $X_2^{(2)}$ sao cho góc tạo bởi U và V là **bé nhất có thể**

  Xét 2 mẫu $(p\ x\ n)X^{(1)}$, $(q\ x\ n)X^{(2)}$ đã được centralize:

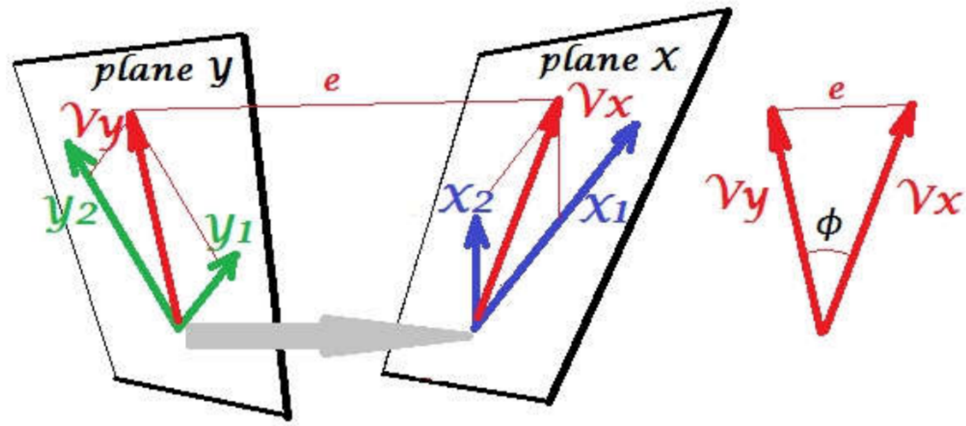
  $\Sigma_{11} = \frac{1}{n}X^{(1)}{X^{(1)}}^T\ (p\ x\ p)$, $\Sigma_{22} = \frac{1}{n}X^{(2)}{X^{(2)}}^T\ (q\ x\ q)$

  $\Sigma_{12} = \frac{1}{n}X^{(1)}{X^{(2)}}^T\ (p\ x\ q)$

$$Corr(U,\ V) = \frac{Cov(U,V)}{\sqrt{Var(U)} * \sqrt{Var(V)}} = \frac{a^T\Sigma_{12}b}{\sqrt{a^T\Sigma_{11}a} * \sqrt{b^T\Sigma_{22}b}}$$

$$= \frac{a^T\frac{1}{n}X^{(1)}{X^{(2)}}^T b}{\sqrt{a^T\frac{1}{n}X^{(1)}{X^{(1)}}^T a} * \sqrt{b^T\frac{1}{n}X^{(2)}{X^{(2)}}^T b}} = \frac{UV^T}{|U||V|}$$

$$là\ góc\ giữa\ vector\ U, V$$

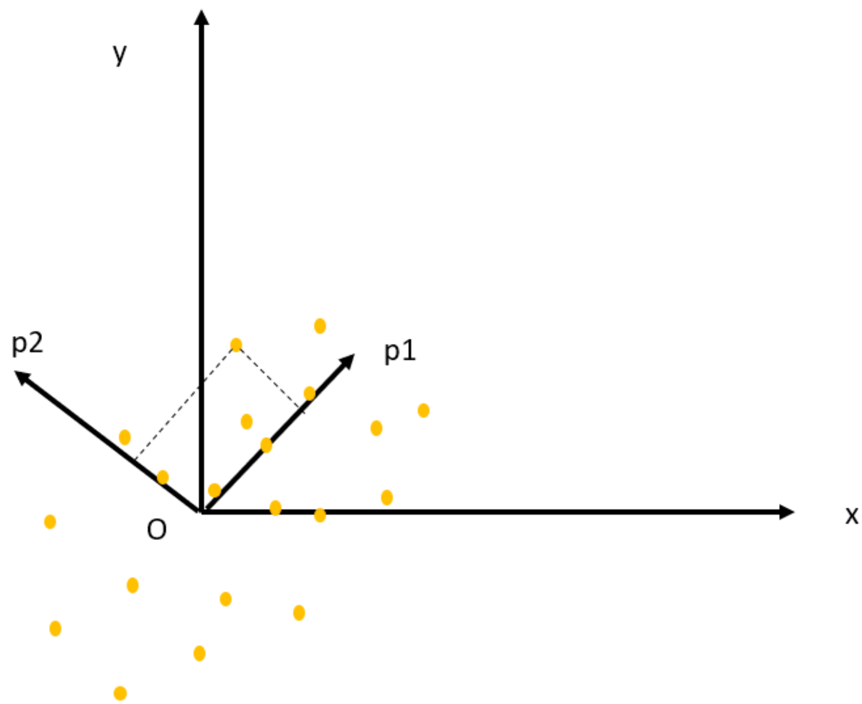


- Về mặt công thức: xét với 2 không gian p chiều và q chiều. Chứa mô tả các biến trong tập $X_1^{(2)}$, $X_2^{(2)}$ ta có thể giải thích dựa trên công thức của CCA.

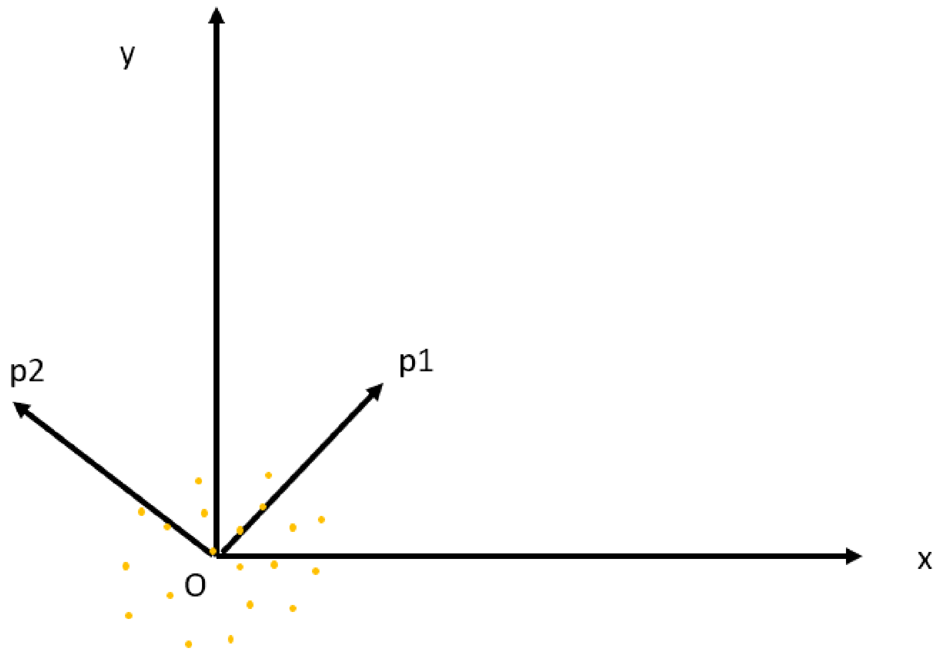
$$U = a^T X^{(1)}$$

$$a = \Sigma_{11}^{-\frac{1}{2}} c = P\Lambda^{-\frac{1}{2}} P^T c \Rightarrow a^T = c^T P \Lambda^{-\frac{1}{2}} P^T$$

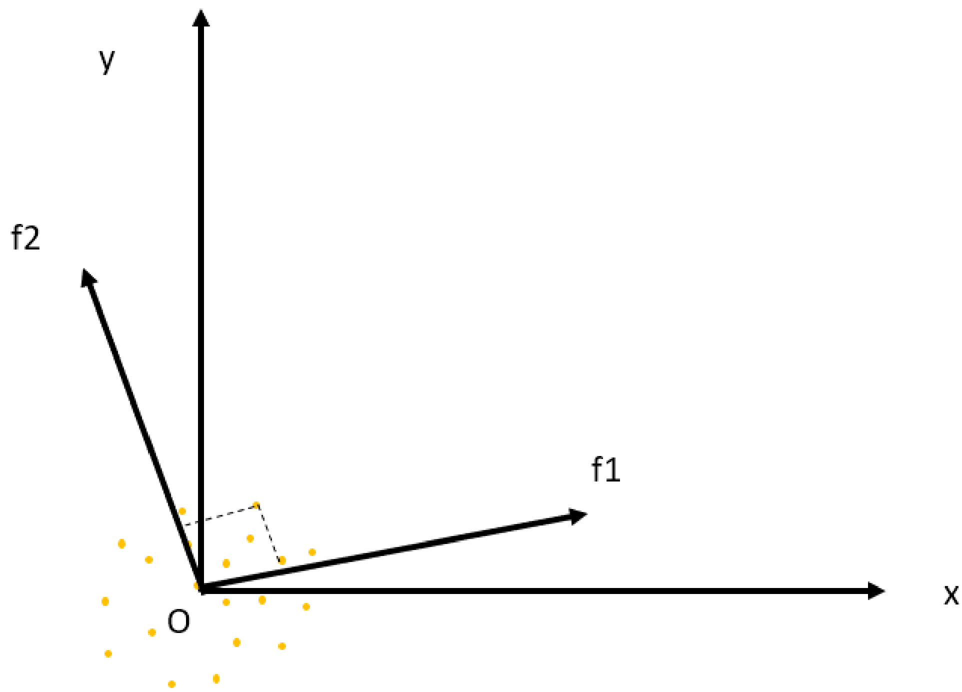
- Với $P = (p1, p2, p3, \ldots)$ là ma trận (p x p) chứa các vector riêng chuẩn hóa (PCA) và $PP^T = I$. Và Λ là ma trận đường chéo chứa các giá trị riêng cũng là phương sai của thành phần chính.
- Xét $P^T X^{(1)}$ chứa các thành phần chính hay là tọa độ của các biến trong một trục tọa độ mới theo pp PCA.



- $\Lambda^{-\frac{1}{2}} P^T X^{(1)}$ với mỗi dòng thứ i sẽ là $\frac{1}{\sqrt{\lambda_i}} p_i X^{(1)}$ đưa về phương sai đơn vị.



- Sau đó t dùng phép xoay về lại trục ban đầu $P\Lambda^{-\frac{1}{2}}P^TX^{(1)}$
- $c^TP\Lambda^{-\frac{1}{2}}P^TX^{(1)}$ Cuối cùng dùng phép chiếu c^T lên vector riêng được xác định từ toàn bộ ma trận covariance của $X^{(1)}, X^{(2)}$ sao cho mối tương quan của U, V là lớn nhất. Tương tự với không gian q chiều



- Tương tự với $V = b^T X^{(2)}$ các bước cũng giống như vậy.
- Mở rộng cho p biến chính tắc giả sử $p \leq q$ trong 2 không gian p chiều và q chiều. Việc phân tích tương quan chính tắc chính là chiếu lên các vector mới sao cho các thành phần chính có tương quan lớn nhất và tương quan lẫn nhau bằng 0…

5. <u>So sánh CCA và PCA:</u>

- Ở PCA - phân tích thành phần chính: ta quan tâm tới việc tối thiểu hóa việc biểu diễn dữ liệu cụ thể hơn với trường hợp tập mẫu $X(n\, x\, p)$ n mẫu và p biến và $p >> n$ khi đó ta sẽ biểu diễn $X = X^{-}(n\, x\, r)$.
- Còn ở CCA - phân tích tương quan chính tắc: ta quan tâm tới việc tổ chức lại dữ liệu và và xác định mối tương quan giữa 2 tập dữ liệu một cách lớn nhất,

trong trường hợp $X^{(1)}(n\,x\,p)$, $X^{(2)}(n\,x\,q)$ và $n >> p, q$ thì cặp vector chính tắc sẽ biểu diễn sự đóng góp vào tổng thể để xác định tương quan.

PCA	CCA
Hàm mục tiêu: Max: $Var(Y_i) = e_1^T \Sigma_{XX} e_1$ Min: $E\|\|x - \hat{x}\|\|^2 = \Sigma e_i^T E(xx^T) e_i$	Hàm mục tiêu: Max: $Corr(U, V) = \frac{a^T \Sigma_{12} b}{\sqrt{a^T \Sigma_{11} a}\sqrt{b^T \Sigma_{22} b}}$

6. Kiểm chứng tính đúng đắn của mô hình (Suy diễn mẫu lớn)

- Vì ta chỉ xét trên tập mẫu hữu hạn, ta cũng phải xem xét việc áp dụng trên cả quần thể mang tính tổng quát hơn.
- Đặt

$$X_j = \begin{bmatrix} X_j^{(1)} \\ \dots\dots \\ X_j^{(2)} \end{bmatrix}, \text{ j = 1,2,...,n}$$

là tập mẫu ngẫu nhiên tuân theo phân phối chuẩn từ quần thể $N_{p+q}(\mu, \Sigma)$ với:

$$\Sigma = \begin{bmatrix} \underset{(pxp)}{\Sigma_{11}} & \vdots & \underset{(pxq)}{\Sigma_{12}} \\ \cdots & \cdots & \cdots \\ \underset{(qxp)}{\Sigma_{21}} & \vdots & \underset{(qxq)}{\Sigma_{21}} \end{bmatrix}$$

- Kiểm định tỷ lệ ước lượng của H_0: $\Sigma_{12} = 0$ với H_1: $\Sigma_{12} \neq 0$ bác bỏ giả thuyết H_0 đối với giá trị lớn:

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax:(08) 38.350.096

$$-2ln(\Lambda) = nln\left(\frac{|S_{11}|\times|S_{22}|}{|S|}\right) = -nln\prod_{i=1}^{p}\left(1-\hat{p}_i^{*2}\right)$$

trong đó

$$S = \begin{bmatrix} S_{11} & S_{11} \\ S_{21} & S_{22} \end{bmatrix}$$

là ước lượng không chênh của Σ. Đối với n lớn, thống kê kiểm định được phân phối xấp xỉ phân phối chi bình phương của 1 biến ngẫu nhiên với bậc tự do pq.

- Λ: nguồn gốc

$$\max_{\mu,\Sigma} L(\mu,\Sigma) = \frac{1}{(2\pi)^{np/2}|\hat{\Sigma}|^{n/2}} e^{-np/2}$$

- Ước lượng cực đại: , $\hat{\Sigma} = S$ (chứng minh ở 4-18, AMSA)
- Ước lượng cực đại theo H0: $\Sigma_{12} = 0 = s_{12} \rightarrow$ $s' = \begin{bmatrix} \frac{S_{11}}{0} & \frac{0}{S_{22}} \end{bmatrix} \rightarrow |s'| = \begin{vmatrix} \frac{S_{11}}{0} & \frac{0}{S_{22}} \end{vmatrix} = |s_{11}||s_{22}|$
- Tỉ lệ ước lượng $\Lambda = \frac{\text{ước lượng cực đại theo } H0}{\text{ước lượng cực đại}} = \left(\frac{|s'|}{|s|}\right)^{-n/2}$
- Để cải thiện xấp xỉ χ^2 phân phối lấy mẫu của $-2ln(\Lambda)$, ta thay n bởi $n - 1 - \frac{1}{2}(p + q + 1)$.
- Vì vậy, đối với n lớn, bác bỏ giả thuyết H_0: $\Sigma_{12} = 0$ với mức ý nghĩa α nếu:

$$-\left(n - 1 - \frac{1}{2}(p + q + 1)\right)ln\prod_{i=1}^{p}\left(1 - \hat{p}_i^{*2}\right) > \chi_{pq}^{2}(x)$$

với $\chi_{pq}^{2}(x)$ là phân vị thứ (100α) (chặn trên) của phân phối chi bình phương với bậc tự do pq.

Nếu bác bỏ H_0: $\Sigma_{12} = 0$, ta xét lần lượt các giả thuyết:
$H_0^{(k)}$: $p_1^*\neq 0,\ p_2^*\neq 0, ..., p_k^*\neq 0, p_{k+1}^* = ... = p_p^* = 0$ với
$H_1^{(k)}$: $p_i^*\neq 0$ với một vài $i \geq k + 1$ (k chạy từ $1 \longrightarrow$ p).

- Bác bỏ $H_0^{(k)}$ tại mức ý nghĩa α nếu:

$$-\left(n - 1 - \frac{1}{2}(p + q + 1)\right)ln\prod_{i=k+1}^{p}\left(1 - \hat{p}_i^{*2}\right) > \chi_{(p-k)(q-k)}^{2}(x)$$

với $\chi_{(p-k)(q-k)}^{2}(x)$ là phân vị thứ (100α) (chặn trên) của phân phối chi bình phương với bậc tự do $(p - k)(q - k)$.

- Kiểm định chuỗi giả thuyết trên cho đến khi có một vài giá trị của k và $H_0^{(k)}$ không bị bác bỏ, điều này thể hiện các độ tương quan còn lại đóng góp không đáng kể trong việc giải thích mối tương quan giữa $X^{(1)}$và $X^{(2)}$.
- Việc kiểm định chuỗi giả thuyết này giúp trong việc chọn ra các cặp biến chính tắc quan trọng (đóng góp đáng kể

trong giải thích mối tương quan giữa $X^{(1)}$ và $X^{(2)}$).

7. Code ứng với giải thuật: Hàm CCA:

```
def CCA(x1,x2):
#   input x1(nxp), x2(nxq)
    #Bước 1
    #tính trung bình
    meanX1 = x1.mean(0)
    meanX2 = x2.mean(0)
    #chuyển vị X1,X2 sang (pxn),(qxn) để dễ xử lý
    x1T = x1.transpose()
    x2T = x2.transpose()
    #tính n,p,q
    n = len(x1)
    p = len(x1[0])
    q = len(x2[0])
    #chuẩn hóa
    for i in range(p):
        dev = math.sqrt(np.cov(x1T[i]))
        for j in range(n):
            x1T[i][j] = float(x1T[i][j] - meanX1[i])/(dev)
    for i in range(q):
        dev = math.sqrt(np.cov(x2T[i]))
        for j in range(n):
            x2T[i][j] = (x2T[i][j] - meanX2[i])/dev
#đặt z1,z2 là biến đã chuẩn hóa
    z1,z2 = x1T,x2T
     #B2: tính ma trận MMT với z1(pxn), z2(qxn) là biến đã chuẩn
hóa và p<=q
    #ma trận xích ma
    cov = np.cov(z1,z2)
    #
    cov11 = cov[:p,:p]
    #
    cov12 = cov[:p,p:]
    #
```

```
    cov21 = cov12.transpose()
    #
    cov22 = cov[p:,p:]
    #
    cov22_1 = np.linalg.inv(cov22)
    #tính căn bậc 2 của ma trận cov11
    sqrt = sqrtm(cov11)
    #lấy nghịch đảo
    sqrtCov11_1 = np.linalg.inv(sqrt)
    #
    mmT = sqrtCov11_1.dot(cov12)
    mmT = mmT.dot(cov22_1)
    mmT = mmT.dot(cov21)
    mmT = mmT.dot(sqrtCov11_1)
#B3,B4: tìm trị riêng và vecto riêng của mmT
    lamda,f = np.linalg.eig(mmT)
    #ma trận hệ số a
    a = sqrtCov11_1.dot(f)
    #ma trận hệ số b
    b_commer = cov22_1.dot(cov21.dot(a))
    bT = b_commer.transpose()
    #chuẩn hóa để phương sai = 1
    for i in range(len(bT)):
        #bT[i] là vecto nằm ngang
        t = bT[i].dot(cov22)
        t = t.dot(bT[i].transpose())
        bT[i] = bT[i]/math.sqrt(t)
    b = bT.transpose()
    #a chuyển vị
    aT = a.transpose()
    #còn tính U,V, hệ số tương quan aT(pxp) x z1T(pxn)
    U = aT.dot(z1)
    V = bT.dot(z2)
    #hệ số tương quan
    corr = np.sqrt(lamda)
    #visualized cặp biến chính tắc thứ nhất
    VisualizedCCAresult(U[0],V[0],aT[0],bT[0],corr[0])
    return a,b,U,V,corr
```

- Hàm VisualizedCCAresult:

```
def VisualizedCCAresult(x,y,a,b,corr):

    # trả về hệ số hồi quy tuyến tính giữa x và y

    beta = estimate_coef(x, y)

     # vẽ đường thẳng hồi quy, và thông báo vecto a,c cùng hệ số
tương quan

    plot_regression_line(x, y,beta,a,b,corr)
```

- Ví dụ:
- Tập dữ liệu được thu thập dựa trên mức độ lo lắng của ba mẹ để dự đoán trọng lượng và "độ lo lắng của trẻ" (có thể được thu thập qua các phản ứng của trẻ như mức độ ăn , hay khóc,..).

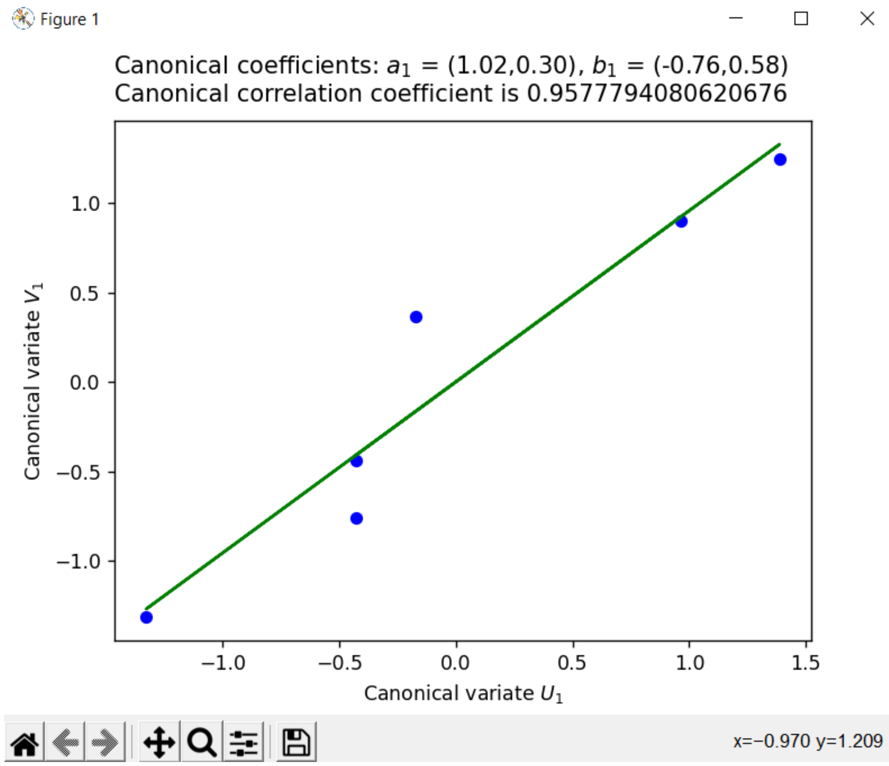
Tập dữ liệu 1:

mức độ lo lắng của mẹ	mức độ lo lắng của cha
3.4	2.9
1.9	8.2
4	3
8.1	2
1.7	1.1
6.1	5.4

Tập dữ liệu 2:

trọng lượng của trẻ	độ lo lắng của trẻ
6.81	8.1
8.23	9.5
6.11	9
5.98	10
9.34	9.2
6.59	10

- Biểu đồ thể hiện sự tương quan của cặp biến chính tắc thứ nhất:



8. Ví dụ

- IOS (Phân tích tương quan chuẩn về sự hài lòng trong công việc) Là một phần của nghiên cứu về ảnh hưởng của cơ cấu tổ chức đối với "sự hài lòng trong công việc", Dunham (4] đã điều tra mức độ mà các thước đo về sự hài lòng trong công việc có liên quan đến đặc điểm công việc. Sử dụng a công cụ khảo sát, Dunham đã thu được các phép đo p = 5 đặc điểm công việc và q = 7 biến sự hài lòng trong công việc đối với n = 784 giám đốc điều hành từ chi nhánh công ty của một tập đoàn bán lẻ lớn.

Các thước đo về sự hài lòng trong công việc có liên quan đến đặc điểm công việc không?

Cho 2 tập dữ liệu X1 và X2, với các biến đặc tính công việc ban đầu là X1,và các biến số về sự hài lòng trong công việc,X2, lần lượt được xác định là:

$$\mathbf{X}^{(1)} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ X_3^{(1)} \\ X_4^{(1)} \\ X_5^{(1)} \end{bmatrix} = \begin{bmatrix} \text{feedback} \\ \text{task significance} \\ \text{task variety} \\ \text{task identity} \\ \text{autonomy} \end{bmatrix}$$

$$\mathbf{X}^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \\ X_3^{(2)} \\ X_4^{(2)} \\ X_5^{(2)} \\ X_6^{(2)} \\ X_7^{(2)} \end{bmatrix} = \begin{bmatrix} \text{supervisor satisfaction} \\ \text{career-future satisfaction} \\ \text{financial satisfaction} \\ \text{workload satisfaction} \\ \text{company identification} \\ \text{kind-of-work-satisfaction} \\ \text{general satisfaction} \end{bmatrix}$$

Các phản hồi cho các biến X (I) và X (2) được ghi lại trên một thang điểm và sau đó được chuẩn hóa. Ma trận tương quan mẫu dựa trên 784 câu trả lời được cho như sau:

$$\mathbf{R} = \left[\begin{array}{c:c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hdashline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right]$$

$$= \left[\begin{array}{ccccc:ccccccc}
1.0 & & & & & .33 & .32 & .20 & .19 & .30 & .37 & .21 \\
.49 & 1.0 & & & & .30 & .21 & .16 & .08 & .27 & .35 & .20 \\
.53 & .57 & 1.0 & & & .31 & .23 & .14 & .07 & .24 & .37 & .18 \\
.49 & .46 & .48 & 1.0 & & .24 & .22 & .12 & .19 & .21 & .29 & .16 \\
.51 & .53 & .57 & .57 & 1.0 & .38 & .32 & .17 & .23 & .32 & .36 & .27 \\
\hdashline
.33 & .30 & .31 & .24 & .38 & 1.0 & & & & & & \\
.32 & .21 & .23 & .22 & .32 & .43 & 1.0 & & & & & \\
.20 & .16 & .14 & .12 & .17 & .27 & .33 & 1.0 & & & & \\
.19 & .08 & .07 & .19 & .23 & .24 & .26 & .25 & 1.0 & & & \\
.30 & .27 & .24 & .21 & .32 & .34 & .54 & .46 & .28 & 1.0 & & \\
.37 & .35 & .37 & .29 & .36 & .37 & .32 & .29 & .30 & .35 & 1.0 & \\
.21 & .20 & .18 & .16 & .27 & .40 & .58 & .45 & .27 & .59 & .31 & 1.0
\end{array}\right]$$

Đặt $U = a^T X^{(1)}, V = b^T X^{(2)}$ Là các tổ hợp tuyến tính tương ứng của $X^{(1)}, X^{(2)}$

Ta xét

$$R_{12} = Corr(X^{(1)}, X^{(2)})$$

$$= \frac{Cov(X^{(1)}, X^{(2)})}{\sqrt{Var(X^{(1)})}\sqrt{Var(X^{(2)})}}$$

Vì dữ liệu đã được chuẩn hoá nên $Var(X^{(2)}) = Var(X^{(1)}) = 1$

$$Nên\ R_{12} = Cov(X^{(1)}, X^{(2)}) = S_{12}$$

Tương tự $R_{21} = S_{21}, R_{11} = S_{11},$ $R_{22} = S_{22}$

Min (p, q) = min (5, 7) = 5 nên ta có 5 biến tương quan chính tắc mẫu và 5 vectơ hệ số chính tắc mẫu được hiển thị qua bảng sau:

	Standardized variables							Standardized variables						
	$z_1^{(1)}$	$z_2^{(1)}$	$z_3^{(1)}$	$z_4^{(1)}$	$z_5^{(1)}$	$\widehat{\rho_1^*}$		$z_1^{(2)}$	$z_2^{(2)}$	$z_3^{(2)}$	$z_4^{(2)}$	$z_5^{(2)}$	$z_6^{(2)}$	$z_7^{(2)}$
$\hat{\mathbf{a}}_1'$:	.42	.21	.17	−.02	.44	.55	$\hat{\mathbf{b}}_1'$:	.42	.22	−.03	.01	.29	.52	−.12
$\hat{\mathbf{a}}_2'$:	−.30	.65	.85	−.29	−.81	.23	$\hat{\mathbf{b}}_2'$:	.03	−.42	.08	−.91	.14	.59	−.02
$\hat{\mathbf{a}}_3'$:	−.86	.47	−.19	−.49	.95	.12	$\hat{\mathbf{b}}_3'$:	.58	−.76	−.41	−.07	.19	−.43	.92
$\hat{\mathbf{a}}_4'$:	.76	−.06	−.12	−1.14	−.25	.08	$\hat{\mathbf{b}}_4'$:	.23	.49	.52	−.47	.34	−.69	−.37
$\hat{\mathbf{a}}_5'$:	.27	1.01	−1.04	.16	.32	.05	$\hat{\mathbf{b}}_5'$:	−.52	−.63	.41	.21	.76	.02	.10

Có thể hiểu rằng

$$U_1 = 0.42z_1^{(1)} + 0.21z_2^{(1)} + 0.17z_3^{(1)} - 0.02z_4^{(1)} + 0.44z_5^{(1)}$$

$$V_1 = 0.42z_1^{(2)} + 0.22 - 0.03z_3^{(2)} + 0.01z_4^{(2)}$$
$$+ 0.29z_5^{(2)} + 0.52z_6^{(2)} - 0.12z_7^{(2)}$$

Tương tự với $U_i, V_i \quad , i = 2, 3, 4, 5$

Các hệ số lớn thể hiện đặc trưng đó ảnh hưởng lớn đến xu hướng của U hay V. ví dụ theo các hệ số, U_1 chủ yếu là biến feedback và Autonnomy, trong khi V_1 đại diện cho supervisor, Career-future và type of work, cùng với company identification.

Để diễn giải mối tương quan cho U_1 và V_1, Cần tính toán mối tương quan mẫu giữa U_1 với các biến thành phần của nó và giữa V_1 với các biến thành phần của nó.

Ta có:

$$Corr(U, X^{(1)}) = aS_{11}D_{11}^{\frac{-1}{2}}$$

$$Corr(V, X^{(2)}) = bS_{22}D_{22}^{\frac{-1}{2}}$$

$$Corr(U, X^{(2)}) = aS_{12}D_{22}^{\frac{-1}{2}}$$

$$Corr(V, X^{(1)}) = bS_{21}D_{11}^{\frac{-1}{2}}$$

Với $D_{11}^{\frac{-1}{2}}$ là ma trận chéo (5x5) với phần tử chéo thứ i là var($x_i^{(1)}$) và $D_{22}^{\frac{-1}{2}}$ là ma trận chéo (7x7) với phần tử chéo thứ i là var($x_i^{(2)}$)

Vì dữ liệu đã được chuẩn hoá nên $D_{11}^{\frac{-1}{2}} = D_{22}^{\frac{-1}{2}} = I$

Ta tính được tương quan mẫu giữa các biến ban đầu và các biến chính tắc:

$\mathbf{X}^{(1)}$ variables	$\hat{U}_1$	$\hat{V}_1$	$\mathbf{X}^{(2)}$ variables	$\hat{U}_1$	$\hat{V}_1$
1. Feedback	.83	.46	1. Supervisor satisfaction	.42	.75
2. Task significance	.74	.41	2. Career-future satisfaction	.35	.65
3. Task variety	.75	.42	3. Financial satisfaction	.21	.39
4. Task identity	.62	.34	4. Workload satisfaction	.21	.37
5. Autonomy	.85	.48	5. Company identification	.36	.65
			6. Kind-of-work satisfaction	.44	.80
			7. General satisfaction	.28	.50

Nhìn thấy 0.83 0.74 0.75 0.62 0.85 lớn nhưng khi nhìn vào cặp vector đầu tiên thấy các hệ số đóng góp không giống nhau…..

Ví dụ 2: (BT 10.12 /591)

Một mẫu ngẫu nhiên gồm n = 70 gia đình sẽ được khảo sát để xác định mối liên hệ giữa các biến "nhân khẩu" và các biến "tiêu dùng" nhất định.

- Tiêu dùng

  $X_1^{(1)}$ = tần suất đi ăn hàng năm tại nhà hàng.

  $X_2^{(1)}$ = tần suất xem phim hàng năm

- Nhân khẩu học

  $X_1^{(2)}$ = tuổi của chủ hộ

  $X_2^{(2)}$ = thu nhập hàng năm của gia đình

  $X_3^{(2)}$ = trình độ văn hóa của chủ hộ

Giả sử 70 quan sát trên các biến trước đó đã được chuẩn hoá và cho ma trận tương quan mẫu:

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right] = \left[\begin{array}{cc|ccc} 1.0 & & & & \\ .80 & 1.0 & & & \\ \hline .26 & .33 & 1.0 & & \\ .67 & .59 & .37 & 1.0 & \\ .34 & .34 & .21 & .35 & 1.0 \end{array}\right]$$

Đặt $U = a^T X^{(1)}$, $V = b^T X^{(2)}$ Là các tổ hợp tuyến tính tương ứng của X(1), X(2)

Min (p, q) = min (2, 3) = 2 nên ta có 2 biến tương quan chính tắc mẫu và 2 vectơ hệ số chính tắc mẫu

Ta xét

$$R_{12} = Corr(X^{(1)}, X^{(2)})$$

$$= \frac{Cov(X^{(1)}, X^{(2)})}{\sqrt{Var(X^{(1)})}\sqrt{Var(X^{(2)})}}$$

$$= \frac{Cov(X^{(1)}, X^{(2)})}{\sqrt{V_{11}}\sqrt{V_{22}}}$$

Vì dữ liệu đã được chuẩn hoá nên
Gọi V_{11} là ma trận ma trận chéo (2x2) với phần tử chéo thứ i là var($x_i^{(1)}$) , V_{22} là ma trận ma trận chéo (3x3) với phần tử chéo thứ i là var($x_i^{(2)}$).
Vì dữ liệu đã được chuẩn hoá nên $V_{11} = V_{22} = I$

$Nên\ R_{12} = Cov(X^{(1)}, X^{(2)}) = S_{12}$

Tương tự $R_{21} = S_{21}$, $R_{11} = S_{11}$, $R_{22} = S_{22}$

$$MM^T = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}$$

$$= \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}^{\frac{-1}{2}} \begin{bmatrix} 0.26 & 0.67 & 0.34 \\ 0.33 & 0.59 & 0.34 \end{bmatrix} \begin{bmatrix} 1.0 & 0.37 & 0.21 \\ 0.37 & 1.0 & 0.35 \\ 0.21 & 0.35 & 1.0 \end{bmatrix}^{-1} \begin{bmatrix} 0.26 & 0.33 \\ 0.67 & 0.59 \\ 0.34 & 0.34 \end{bmatrix} \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}^{\frac{-1}{2}}$$

$= \begin{bmatrix} 0.322 & 0.208 \\ 0.208 & 0.186 \end{bmatrix}$

Ta tìm 2 giá trị riêng của ma trận MM^T:

$$det\left(MM^T - \lambda I\right) = 0$$

$$\Leftrightarrow \lambda^2 - 0.508\lambda + 0.017 = 0$$

Ta được:

$$\lambda_1 = 0.473$$

$$\lambda_2 = 0.035$$

Vậy độ tương quan chính tắc:

Corr(U_1, V_1) $= \sqrt{\lambda_1}$=0.69

Corr(U_2, V_2) $= \sqrt{\lambda_2}$=0.19

Vector riêng f_1 ứng giá trị riêng λ_1 là $\begin{bmatrix} 1.378 \\ 1 \end{bmatrix}$

f_1 sau khi Scale: $\begin{bmatrix} 0.81 \\ 0.587 \end{bmatrix}$

Vector riêng f_2 ứng giá trị riêng λ_2 là $\begin{bmatrix} -0.725 \\ 1 \end{bmatrix}$

f_2 sau khi Scale: $\begin{bmatrix} -0.587 \\ 0.81 \end{bmatrix}$

<u>Vector hệ số a:</u>

$a_1 = S_{11}^{-\frac{1}{2}} f_1$

$= \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}^{\frac{-1}{2}} \begin{bmatrix} 0.81 \\ 0.587 \end{bmatrix}$

$$= \begin{bmatrix} 0.77 \\ 0.271 \end{bmatrix}$$

$$a_2 = S_{11}^{-\frac{1}{2}} f_2$$

$$= \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} -0.587 \\ 0.81 \end{bmatrix}$$

$$= \begin{bmatrix} -1.479 \\ 1.645 \end{bmatrix}$$

Hay a= $\begin{bmatrix} 0.77 & 0.271 \\ -1.479 & 1.645 \end{bmatrix}$

<u>Vector hệ số b:</u>

Đặt

$$b' = S_{22}^{-1} . S_{21} . a$$

$$= \begin{bmatrix} 1.0 & 0.37 & 0.21 \\ 0.37 & 1.0 & 0.35 \\ 0.21 & 0.35 & 1.0 \end{bmatrix}^{-1} \begin{bmatrix} 0.26 & 0.33 \\ 0.67 & 0.59 \\ 0.34 & 0.34 \end{bmatrix} \begin{bmatrix} 0.77 & 0.271 \\ -1.479 & 1.645 \end{bmatrix}$$

$$t = b'^T S_{22} . b'$$

Ta tính $b = \frac{1}{\sqrt{t}} . b'$ (Chúng ta cần scale b để Var(V)=1)

$$b1' = [0.034,\ 0.618,\ 0.131]^T$$

$$t1 = 0.473$$

$$b1 = [0.05,\ 0.9,\ 0.19]^T$$

$$b2' = [0.187, -0.109, 0.055]^T$$

$$t2 = 0.035$$

$$b2 = [1 \quad -0.58 \quad 0.29]^T$$

Hay

$$b = \begin{bmatrix} 0.05 & 0.9 & 0.19 \\ 1 & -0.58 & 0.29 \end{bmatrix}$$

Có thể hiểu

$U_1 = 0.77x_1^{(1)} + 0.271x_2^{(1)}$

$V_1 = 0.05x_1^{(2)} + 0.9x_2^{(2)} + 0.19x_3^{(2)}$ (tương tự với U_2 và V_2)

Ta có thể xem hệ số tương quan như là một "Hệ số ảnh hưởng" . Nếu hệ số ảnh hưởng càng cao thể hiện đặc trưng đó ảnh hưởng càng lớn đến xu hướng của U hay V. Ví dụ theo các hệ số, U_1 chủ yếu bị tác động bởi biến tần suất xem phim phim hằng năm , trong khi V_1 đại diện cho tuổi của chủ hộ và thu nhập hằng năm của gia đình.

Để diễn giải mối tương quan cho U_1 và V_1, Cần tính toán mối tương quan mẫu giữa U_1 với các biến thành phần của nó và giữa V_1 với các biến thành phần của nó.

$$Corr(U, X^{(2)}) = \frac{Cov(U, X^{(2)})}{\sqrt{Var(U)}\sqrt{Var(X^{(2)})}}$$

Mà Var(U)=1

=> $R_{12} = \frac{Cov(U, X^{(2)})}{\sqrt{Var(X^{(2)})}}$

$$= (D_2)^{\frac{-1}{2}} Cov(U, X^{(2)})$$

Vì dữ liệu đã được chuẩn hoá nên $D_2 = I$ với D_2 là ma trận chéo với phần tử chéo thứ i là $x_1^{(2)}$

Vậy $Corr(U, X^{(2)}) = Cov(U, X^{(2)})$

$$= Cov(aX^{(1)}, X^{(2)})$$

$$= aCov(X^{(1)}, X^{(2)})$$

$$= aS_{12}$$

Vậy $Corr(U, X^{(2)}) = aS_{12}$

$$= \begin{bmatrix} 0.77 & 0.271 \\ -1.479 & 1.645 \end{bmatrix} \begin{bmatrix} 0.26 & 0.67 & 0.34 \\ 0.33 & 0.59 & 0.34 \end{bmatrix}$$

$$= \begin{bmatrix} 0.29 & 0.676 & 0.354 \\ 0.158 & -0.02 & 0.056 \end{bmatrix}$$

Tương tự

$$Corr(U, X^{(1)}) = aS_{11}$$

$$= \begin{bmatrix} 0.77 & 0.271 \\ -1.479 & 1.645 \end{bmatrix} \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.987 & 0.887 \\ -0.163 & 0.462 \end{bmatrix}$$

$$Corr(V, X^{(1)}) = bS_{21}$$

$$= \begin{bmatrix} 0.05 & 0.9 & 0.19 \\ 1 & -0.58 & 0.29 \end{bmatrix} \begin{bmatrix} 0.26 & 0.33 \\ 0.67 & 0.59 \\ 0.34 & 0.34 \end{bmatrix}$$

$$= \begin{bmatrix} 0.681 & 0.612 \\ -0.03 & 0.086 \end{bmatrix}$$

$$Corr(V, X^{(2)}) = bS_{22}$$

$$= \begin{bmatrix} 0.05 & 0.9 & 0.19 \\ 1 & -0.58 & 0.29 \end{bmatrix} \begin{bmatrix} 1.0 & 0.37 & 0.21 \\ 0.37 & 1.0 & 0.35 \\ 0.21 & 0.35 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.423 & 0.985 & 0.516 \\ 0.846 & -0.109 & 0.297 \end{bmatrix}$$

Hay :

$X^{(1)}$	U_2	V_2	U_1	V_1
Tần suất đi ăn hàng năm tại nhà hàng.	-0.163	-0.03	0.987	0.681
Tần suất xem phim hàng năm.	0.462	0.086	0.887	0.612

$X^{(2)}$	U_2	V_2	U_1	V_1
Tuổi của chủ hộ.	0.158	0.846	0.290	0.423
Thu nhập hàng năm của gia đình.	-0.02	-0.109	0.676	0.985
Trình độ văn hóa của chủ hộ.	0.056	0.297	0.354	0.516

Ta có thể thấy mối tương quan giữa V_1 và $X^{(2)}$ là lớn (0.985) điều này chỉ ra rằng khi thu nhập hằng năm của hộ gia đình cao thì họ có khả năng chi tiêu nhiều hơn. Nếu mối tương quan này được lặp lại ở một hay nhiều nhóm đối tượng khác, có thể sử dụng mức thu nhập của gia đình để tiên đoán mức độ chi tiêu của hộ gia đình đó.

Hay mối tương quan giữa U_2 và $X^{(1)}$ là nhỏ (= -0.163) có thể thấy được tần số đi ăn tại nhà hàng hằng năm của gia đình đó không liên quan gì đến tuổi, thu nhập hay trình độ văn hoá của chủ hộ.

Một điều cực kỳ quan trọng cần nằm lòng khi diễn dịch kết quả phân tích tương quan là hệ số tương quan không hẳn phản ảnh mối quan hệ nhân quả (cause-and-effect relationship). Nếu không có lý do sinh học, không thể và không nên diễn dịch hệ số tương quan theo định hướng nguyên nhân và hệ quả. Chẳng hạn mối tương quan giữa thu nhập hằng năm của gia đình với các biến chi tiêu khá cao nhưng điều này không có nghĩa là thu nhập hằng năm cao là nguyên nhân chi tiêu nhiều hay ngược lại. Mối liên hệ mà chúng ta quan sát chỉ đơn thuần là tương quan.Việc sử dụng và diễn dịch mối tương quan đó trong bối cảnh ra sao còn tùy thuộc vào kinh nghiệm và kiến thức của từng bộ môn khoa học.

B. Ứng dụng

1. **<u>Phát biểu bài toán</u>**

Input: Ảnh của một kí số viết tay

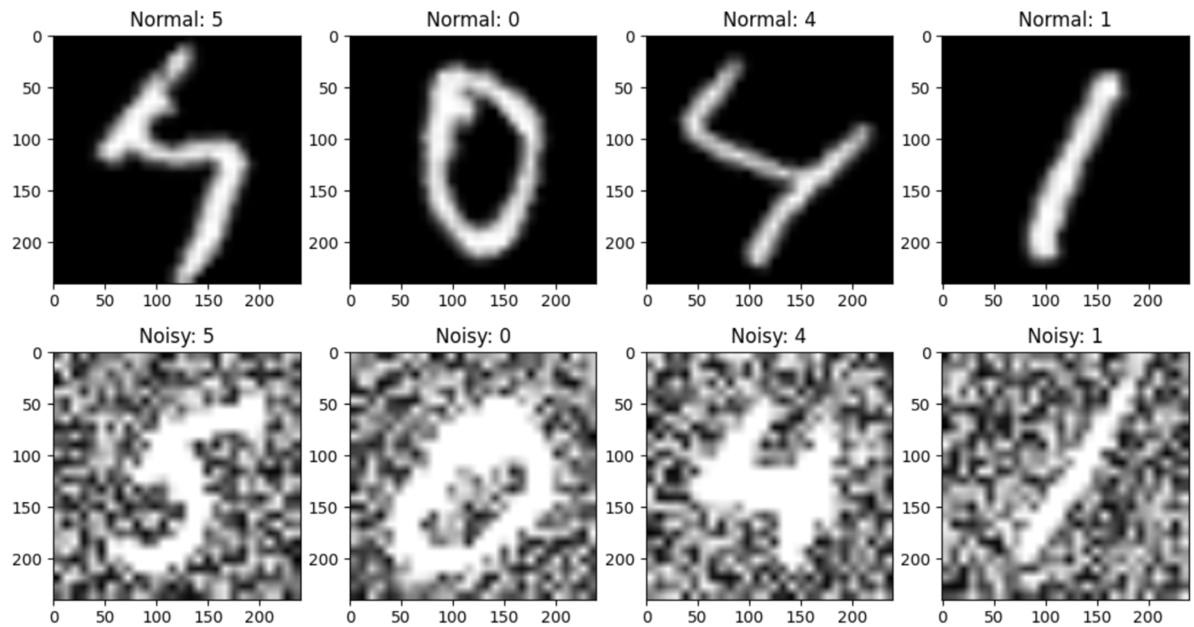
Output: Nhãn của ảnh tương ứng (từ 0 đến 9)

2. **<u>Phương pháp tổng quát</u>**

Đồ án sử dụng tập dữ liệu MNIST (Modified National Institute of Standards and Technology) Bộ dữ liệu MNIST được chia thành hai phần: Dữ liệu dành cho quá trình huấn luyện và dữ liệu dành cho quá trình kiểm tra.

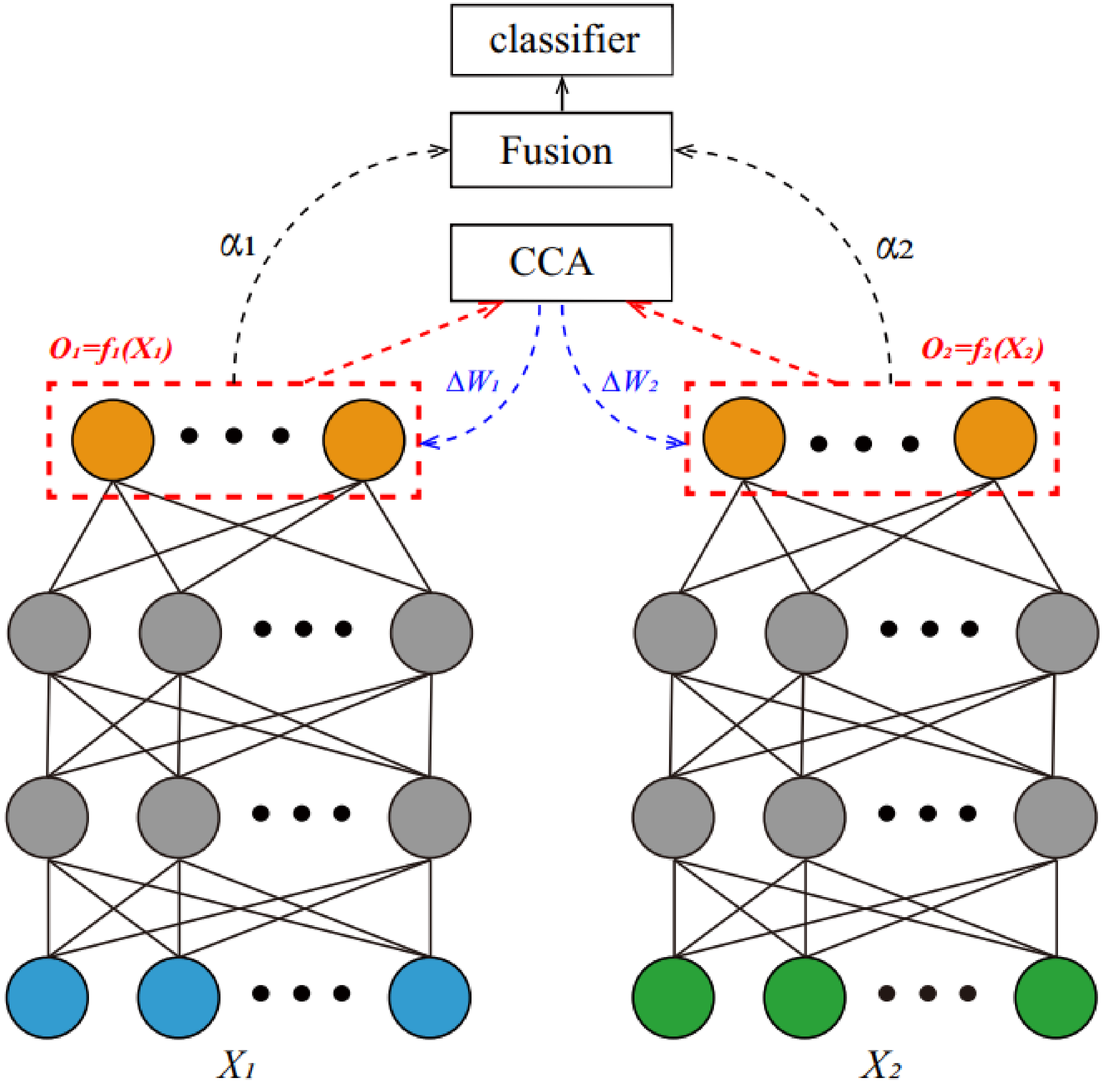
- Dữ liệu huấn luyện gồm 50000 ảnh
- Dữ liệu kiểm tra gồm 10000 ảnh.

Tất cả đều là hình ảnh đen trắng các chữ số viết tay từ 0 tới 9 có kích thước 28 pixel x 28 pixel đã được gắn nhãn đúng. Một số hình ảnh ngẫu nhiên được trích xuất từ bộ dữ liệu MNIST:



Để phục vụ cho mô hình mạng học sâu, với mỗi mẫu dữ liệu ta có 2 view dữ liệu:

- View1: Dữ liệu gốc ban đầu
- View2: Dữ liệu bị làm nhiễu



Gồm 2 công đoạn. Ở công đoạn thứ nhất, sẽ cố gắng rút trích đặc trưng sao cho khi áp dụng phương pháp CCA thì tương quan của đặc trưng trong không gian mới là lớn nhất, với mong muốn sẽ bất biến với nhiễu. Và trong giai đoạn suy diễn (inference stage) ta sẽ phân lớp dựa vào sub network của View1.

3. Phương pháp thực hiện

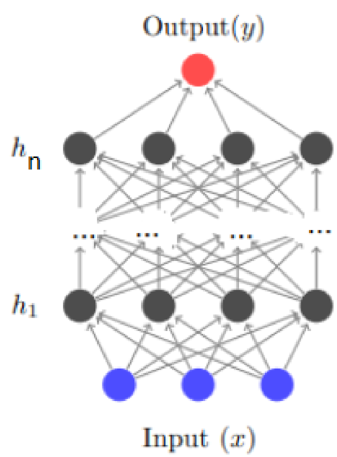
3.1. Giới thiệu DCCA

3.1.1. Động lực nghiên cứu

- Theo các kết quả thực nghiệm thì DCCA mang lại các thể hiện tương quan cao hơn KCCA hay CCA truyền thống
- DCCA kết hợp trong các mô hình học máy giúp nâng cao hiệu quả của các mô hình.

3.1.2. Nhắc lại học sâu

Học sâu là một nhánh của học máy sử dụng nhiều lớp Neurol network để đưa ra một mô hình toán học trên dữ liệu có sẵn.



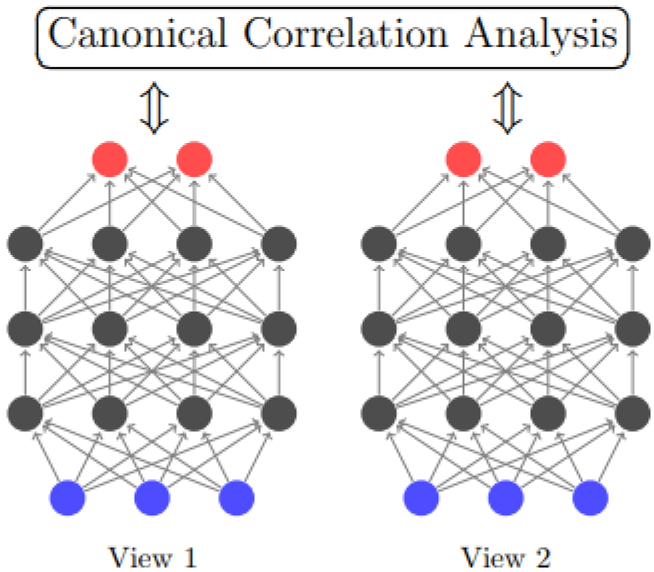
$h_1 = s(W_1x + b_1)$

$$h_2 = s(W_2h_1 + b_2), \dots, h_n = s(W_dh_{d-1} + b_d)$$

Với s là hàm phi tuyến

3.1.3. Cấu trúc DCCA

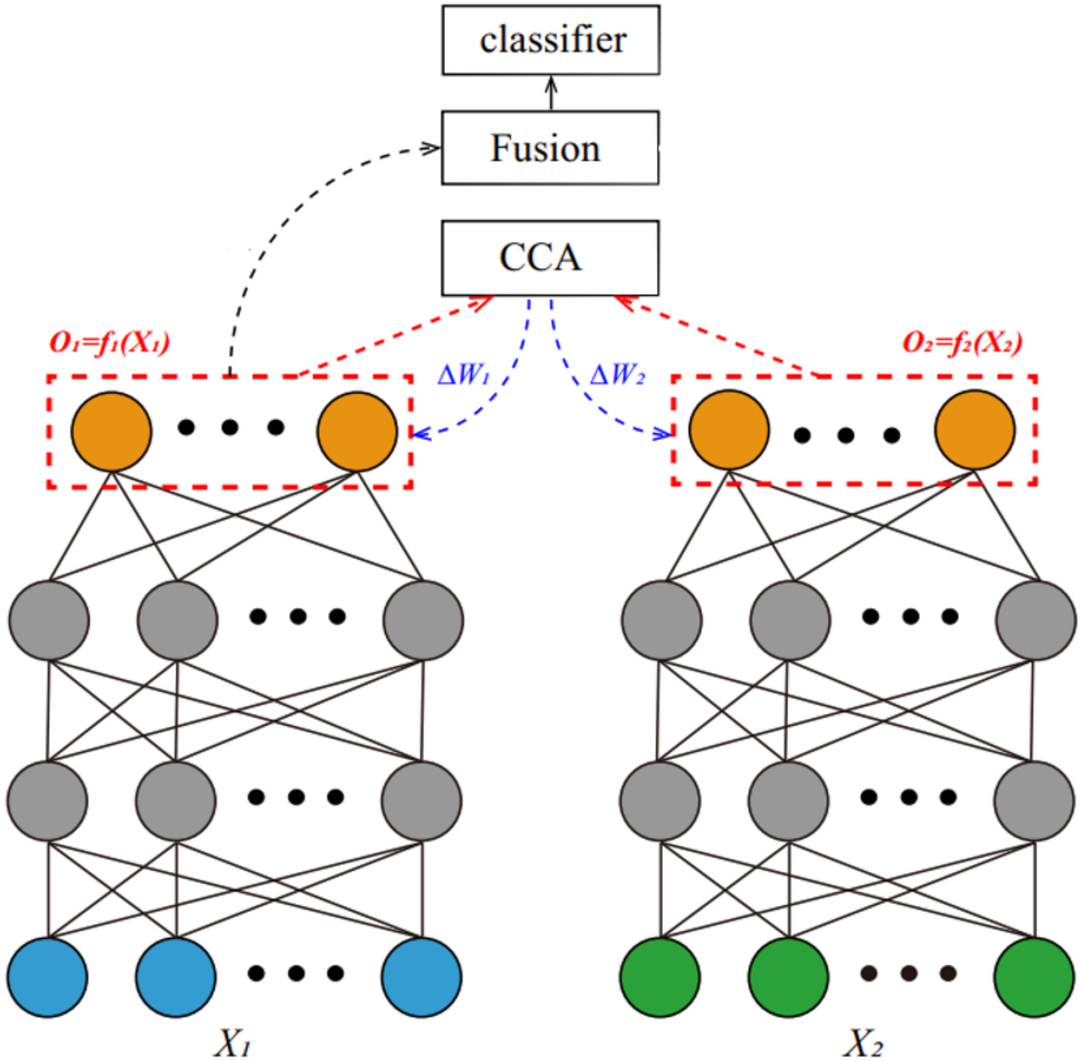
Khác với CCA là một kỹ thuật cổ điển tìm các cặp biến không quan sát được dựa vào 2 tập biến được quan sát sao cho tương quan của chúng là lớn nhất, DCCA học các biến đổi phi tuyến phức tạp để khám phá ra các mối quan hệ tiềm ẩn trong 2 view dữ liệu thông qua đặc trưng được chiết xuất trong 2 khung hình khác nhau (ở đây là ảnh ban đầu và ảnh nhiễu).



DCCA học các phép biến đổi phi tuyến qua 2 sub-network từ dữ liệu ở 2 view khác nhau sao cho các đặc trưng thu được có tương quan cao. Các tham số của 2 sub-network được học cùng một lúc để tối đa hoá mối tương quan của các đặc trưng sau khi rút trích từ 2 sub network.

3.2. Kiến trúc mô hình

3.2.1. Ứng dụng Deep CCA trong mô hình



Trong mô hình trên:

Dữ liệu $X_1 \in R^{n_1 \times d_1}$ và $X_2 \in R^{n_2 \times d_2}$ trong từng view được chuyển đổi bởi một mạng nơron riêng biệt tương ứng. Ta cập nhật các tham số của mạng dựa vào hàm lỗi CCA sẽ được trình bày bên dưới. Từ đó ta có được features là kết quả đã được huấn luyện là những đặc trưng quan trọng (hay có đóng góp nhiều nhất). Các feature này sau đó được đưa vào một method classification cụ thể để thực hiện phân lớp.

Gọi $X_1 \in R^{n_1 \times d_1}$ là là ma trận dữ liệu từ View1, $X_2 \in R^{n_2 \times d_2}$ là là ma trận dữ liệu từ View2.

Để Có được features từ dữ liệu ban đầu, ta xây dựng hai mạng noron với

$$H_2^1 = f_1^1(X_1^1, W_1^1) = s\left(\left(W_1^1\right)^T X_1 + bias\right),$$

$$H_2^2 = f_1^2(X_1^2, W_1^2) = s\left(\left(W_1^2\right)^T X_1^2 + bias\right)$$

$$\Rightarrow H_i^k = f_{i-1}^k\left(X_{i-1}^k, W_{i-1}^k\right) = s\left(\left(W_{i-1}^k\right)^T X_{i-1}^k + bias\right), \quad k = 1, 2$$

Với trọng số $W^1 \in R^{o \times o}, W^2 \in R^{o \times o}$ biểu thị tất cả các tham số cho phép biến đổi phi tuyến tính và $H_1 \in R^{o \times m}$, $H_2 \in R^{o \times m}$ là kết quả của mạng noron k lớp.

Mục tiêu của DCCA là cùng tìm các tham số W_i^1, W_i^2 sao cho mối tương quan giữa H_n^1 và H_n^2 - out features sau khi qua CCA là cao nhất.

Hay một cách tổng quát Mục tiêu của DCCA là cùng tìm các tham số W_i^k, $với\, i = 1, \ldots, n;\ k = 1, 2$ để mối tương quan của categories trong feature trong không gian mới sau khi qua CCA là cao nhất

Lưu ý: với Deep CCA điều chỉnh tham số dựa vào hàm lỗi có chứa hệ số tương quan cần tính dựa trên một tập dữ liệu, vì thế không thể áp dụng SGD bình thường mà phải áp dụng cho full batch.

- Các bước train một DCCA model
  1. Chia nhỏ tập train ra thành nhiều batch.
  2. Train mỗi tập tập dữ liệu ban đầu qua nhiều lớp của mạng sâu.
  3. Cùng cập nhật các tham số W_i^1, W_i^2 ở 2 sub-network dựa trên hàm lỗi CCA sẽ được trình bày bên dưới.

Với từng output cho từng mạng tương ứng thực hiện centralize H_1, H_2:

$\underline{H}_1 = H_1 - \frac{1}{m}H_1$, $\underline{H}_2 = H_2 - \frac{1}{m}H_2$

Ma trận hiệp phương sai $\hat{\Sigma}_{12} = \frac{1}{m-1}\left(\underline{H}_1 \underline{H}_2^T\right)$,

$$\hat{\Sigma}_{11} = \frac{1}{m-1}\left(\underline{H}_1 \underline{H}_1^T\right) + r_1 I$$

$$\hat{\Sigma}_{22} = \frac{1}{m-1}\left(\underline{H}_2 \underline{H}_2^T\right) + r_2 I$$

Đặt $T = \hat{\Sigma}_{11}^{-1/2} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1/2}$

SVD của ma trận T: $T = UDV^T$

Quay lại bài toán CCA đã được chứng minh ở phần trước,

$Corr(U^T H_{1,} V^T H_2) = \sqrt{\lambda}$, Với λ là giá trị riêng lớn nhất của ma trận TT^T

- **Hàm lỗi DCCA**

  Vậy $f = Corr(U^T H_1, V^T H_2) = tr(TT^T)^{1/2}$

  Với mục tiêu: $w_1, w_2 = arg\,min_{w_{1,} w_2}\ (-\ Corr(U^T H_1, V^T H_2))$

  $$\frac{\partial corr(U^T H_1, V^T H_2)}{\partial H_1} = \frac{1}{m-1}(2\Lambda_{11}\underline{H_1} + \Lambda_{12}\underline{H_2})$$

  Với $\Lambda_{12} = \hat{\Sigma}_{11}^{-1/2} UV^T \hat{\Sigma}_{22}^{-1/2}$ là đạo hàm riêng của f theo $\hat{\Sigma}_{12}$

  và $\Lambda_{11} = \frac{-1}{2}\hat{\Sigma}_{11}^{-1/2} UDU^T \hat{\Sigma}_{22}^{-1/2}$ là đạo hàm riêng của f theo $\hat{\Sigma}_{11}$.

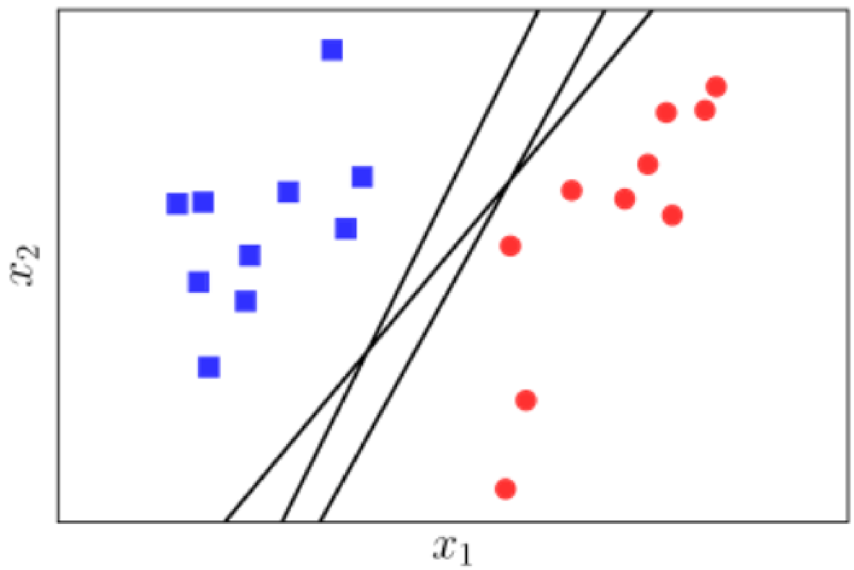
3.2.2. SVM classification

Sau khi có được những đặc trưng quan trọng từ DCCA ta dùng SVM classification để thực hiện phân lớp.

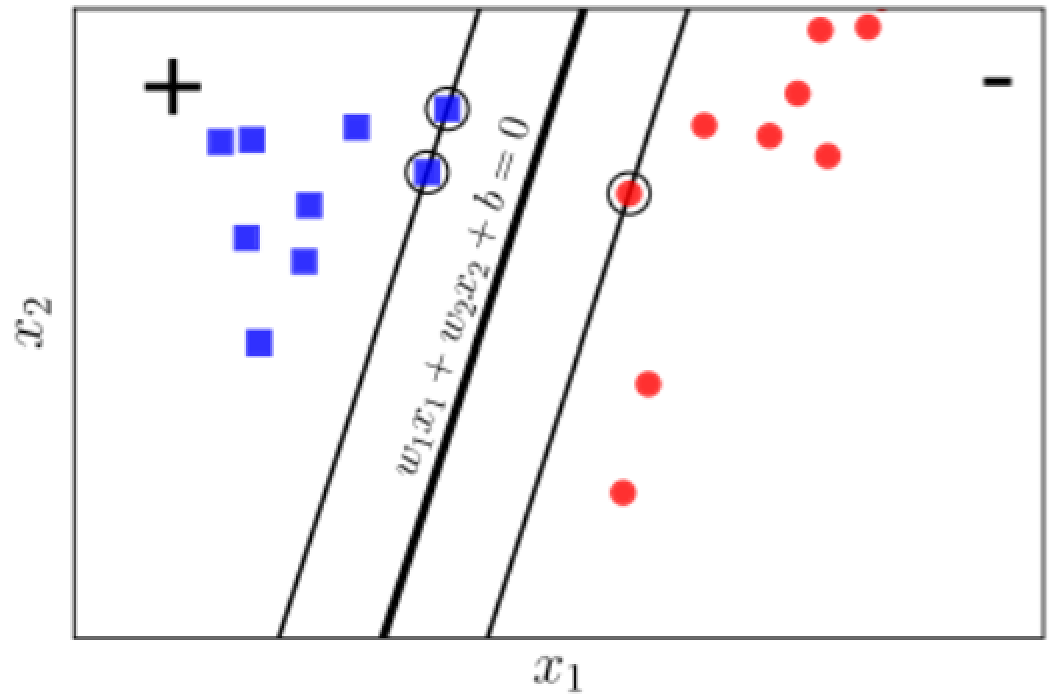
SVM (Support Vector Machine) là một thuật toán học máy có giám sát được sử dụng rất phổ biến ngày nay trong các bài toán phân lớp (classification) hay hồi qui (Regression).

Ý tưởng của SVM là tìm một siêu phẳng (hyper lane) để phân tách các điểm dữ liệu. Siêu phẳng này sẽ chia không gian thành các miền khác nhau và mỗi miền sẽ chứa một loại giữ liệu.

Vấn đề là có rất nhiều siêu phẳng, chúng ta phải chọn cái nào để tối ưu nhất ?



Giả sử chúng ta phải phân loại tập dữ liệu các lớp dương (màu xanh) nhãn là 1 và các dữ liệu lớp âm (màu đỏ) nhãn là -1 và mặt phẳng $W^T x + b = W_1 x_1 + W_2 x_2 + b = 0$ là mặt phân chia giữa 2 lớp.



Vậy với cặp dữ liệu (x_n, y_n) bất kỳ khoảng cách từ điểm đó tới mặt phân chia là

$$\frac{y_n(W^T x_n + b)}{||W||_2}$$

Với mặt phần chia như trên, *margin* được tính là khoảng cách gần nhất từ 1 điểm tới mặt đó (bất kể điểm nào trong hai classes):

$$margin = min_n \frac{y_n(W^T x_n + b)}{||W||_2}$$

Bài toán tối ưu trong SVM chính là bài toán tìm w và b sao cho *margin* này đạt giá trị lớn nhất:

$$(W, b) = argmax_{w,b}(min_n \frac{y_n(W^T x_n + b)}{||W||_2})$$

$$= argmax_{w,b}\left(\frac{1}{||W||_2} min_n y_n (W^T x_n + b)\right)$$

4. Ứng dụng mô hình (code)

Hàm lỗi CCA dùng để train mô hình:

```
    def loss(self, H1, H2):
        """

         It is the loss function of CCA as introduced in the original
paper. There can be other formulations.

        """

        r1 = 1e-3
        r2 = 1e-3
        eps = 1e-9

        H1, H2 = H1.t(), H2.t()
        # assert torch.isnan(H1).sum().item() == 0
        # assert torch.isnan(H2).sum().item() == 0

        o1 = o2 = H1.size(0)

        m = H1.size(1)
#         print(H1.size())

        H1bar = H1 - H1.mean(dim=1).unsqueeze(dim=1)
        H2bar = H2 - H2.mean(dim=1).unsqueeze(dim=1)
        # assert torch.isnan(H1bar).sum().item() == 0
        # assert torch.isnan(H2bar).sum().item() == 0

        SigmaHat12 = (1.0 / (m - 1)) * torch.matmul(H1bar, H2bar.t())
        SigmaHat11 = (1.0 / (m - 1)) * torch.matmul(H1bar,
                                                    H1bar.t()) + r1 *
torch.eye(o1, device=self.device)
```

```
        SigmaHat22 = (1.0 / (m - 1)) * torch.matmul(H2bar,
                                                     H2bar.t()) + r2 *
torch.eye(o2, device=self.device)
        # assert torch.isnan(SigmaHat11).sum().item() == 0
        # assert torch.isnan(SigmaHat12).sum().item() == 0
        # assert torch.isnan(SigmaHat22).sum().item() == 0

           # Calculating the root inverse of covariance matrices by
using eigen decomposition
        [D1, V1] = torch.symeig(SigmaHat11, eigenvectors=True)
        [D2, V2] = torch.symeig(SigmaHat22, eigenvectors=True)
        # assert torch.isnan(D1).sum().item() == 0
        # assert torch.isnan(D2).sum().item() == 0
        # assert torch.isnan(V1).sum().item() == 0
        # assert torch.isnan(V2).sum().item() == 0

        # Added to increase stability
        posInd1 = torch.gt(D1, eps).nonzero()[:, 0]
        D1 = D1[posInd1]
        V1 = V1[:, posInd1]
        posInd2 = torch.gt(D2, eps).nonzero()[:, 0]
        D2 = D2[posInd2]
        V2 = V2[:, posInd2]
        # print(posInd1.size())
        # print(posInd2.size())

        SigmaHat11RootInv = torch.matmul(
            torch.matmul(V1, torch.diag(D1 ** -0.5)), V1.t())
        SigmaHat22RootInv = torch.matmul(
            torch.matmul(V2, torch.diag(D2 ** -0.5)), V2.t())

        Tval = torch.matmul(torch.matmul(SigmaHat11RootInv,
                                                          SigmaHat12),
SigmaHat22RootInv)
#         print(Tval.size())

        if self.use_all_singular_values:
                   # all singular values are used to calculate the
correlation
```

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax:(08) 38.350.096

```
            tmp = torch.matmul(Tval.t(), Tval)
            corr = torch.trace(torch.sqrt(tmp))
            # assert torch.isnan(corr).item() == 0
        else:
            # just the top self.outdim_size singular values are used
            trace_TT = torch.matmul(Tval.t(), Tval)
                                        trace_TT    =    torch.add(trace_TT,
(torch.eye(trace_TT.shape[0])*r1).to(self.device))   #   regularization
for more stability
            U, V = torch.symeig(trace_TT, eigenvectors=True)
                                          U    =    torch.where(U>eps,    U,
(torch.ones(U.shape).double()*eps).to(self.device))
            U = U.topk(self.outdim_size)[0]
            corr = torch.sum(torch.sqrt(U))
        return -corr
```

Cấu trúc của mạng neural network:

```
layer_sizes1 = [1024, 1024, 1024, outdim_size]
class MlpNet(nn.Module):
    def __init__(self, layer_sizes, input_size):
        super(MlpNet, self).__init__()
        layers = []
        layer_sizes = [input_size] + layer_sizes
        for l_id in range(len(layer_sizes) - 1):
            if l_id == len(layer_sizes) - 2:
                layers.append(nn.Sequential(
                       nn.BatchNorm1d(num_features=layer_sizes[l_id],
affine=False),
                      nn.Linear(layer_sizes[l_id], layer_sizes[l_id +
1]),
                ))
            else:
                layers.append(nn.Sequential(
                      nn.Linear(layer_sizes[l_id], layer_sizes[l_id +
1]),
                    nn.Sigmoid(),
```

```
                        nn.BatchNorm1d(num_features=layer_sizes[l_id +
1], affine=False),
                ))
        self.layers = nn.ModuleList(layers)

    def forward(self, x):
        for layer in self.layers:
            x = layer(x)
        return x
```

Train DCCA:

```
def  fit(self,  x1,  x2,  vx1=None,  vx2=None,  tx1=None,  tx2=None,
checkpoint='checkpoint.model'):
        """

        x1, x2 are the vectors needs to be make correlated
        dim=[batch_size, feats]

        """
        x1.to(self.device)
        x2.to(self.device)

        data_size = x1.size(0)

        if vx1 is not None and vx2 is not None:
            best_val_loss = 0
            vx1.to(self.device)
            vx2.to(self.device)
        if tx1 is not None and tx2 is not None:
            tx1.to(self.device)
            tx2.to(self.device)

        train_losses = []
        for epoch in range(self.epoch_num):
            epoch_start_time = time.time()
            self.model.train()
```

```
            # tạo ra list chứa idx random 2 chiều
            batch_idxs = list(BatchSampler(RandomSampler(
                        range(data_size)), batch_size=self.batch_size,
drop_last=False))
            for batch_idx in batch_idxs:
                self.optimizer.zero_grad()
                batch_x1 = x1[batch_idx, :]
                batch_x2 = x2[batch_idx, :]
                o1, o2 = self.model(batch_x1, batch_x2)
                loss = self.loss(o1, o2)
                train_losses.append(loss.item())
                loss.backward()
                self.optimizer.step()
            train_loss = np.mean(train_losses)

                   info_string = "Epoch {:d}/{:d} - time: {:.2f} -
training_loss: {:.4f}"
            if vx1 is not None and vx2 is not None:
                with torch.no_grad():
                    self.model.eval()
                    val_loss = self.test(vx1, vx2)
                                      info_string += " - val_loss:
{:.4f}".format(val_loss)
                    if val_loss < best_val_loss:
                        self.logger.info(
                                  "Epoch {:d}: val_loss improved from
{:.4f} to {:.4f}, saving model to {}".format(epoch + 1,
best_val_loss, val_loss, checkpoint))
                        best_val_loss = val_loss
                                  torch.save(self.model.state_dict(),
checkpoint)
                    else:
                           self.logger.info("Epoch {:d}: val_loss did
not improve from {:.4f}".format(
                            epoch + 1, best_val_loss))
            else:
                torch.save(self.model.state_dict(), checkpoint)
            epoch_time = time.time() - epoch_start_time
            self.logger.info(info_string.format(
```

```
                epoch + 1, self.epoch_num, epoch_time, train_loss))
        # train_linear_cca
        if self.linear_cca is not None:
            _, outputs = self._get_outputs(x1, x2)
            self.train_linear_cca(outputs[0], outputs[1])

        checkpoint_ = torch.load(checkpoint)
        self.model.load_state_dict(checkpoint_)
        if vx1 is not None and vx2 is not None:
            loss = self.test(vx1, vx2)
                         self.logger.info("loss  on  validation  data:
{:.4f}".format(loss))

        if tx1 is not None and tx2 is not None:
            loss = self.test(tx1, tx2)
                              self.logger.info('loss   on   test   data:
{:.4f}'.format(loss))
```

Train SVM classification:

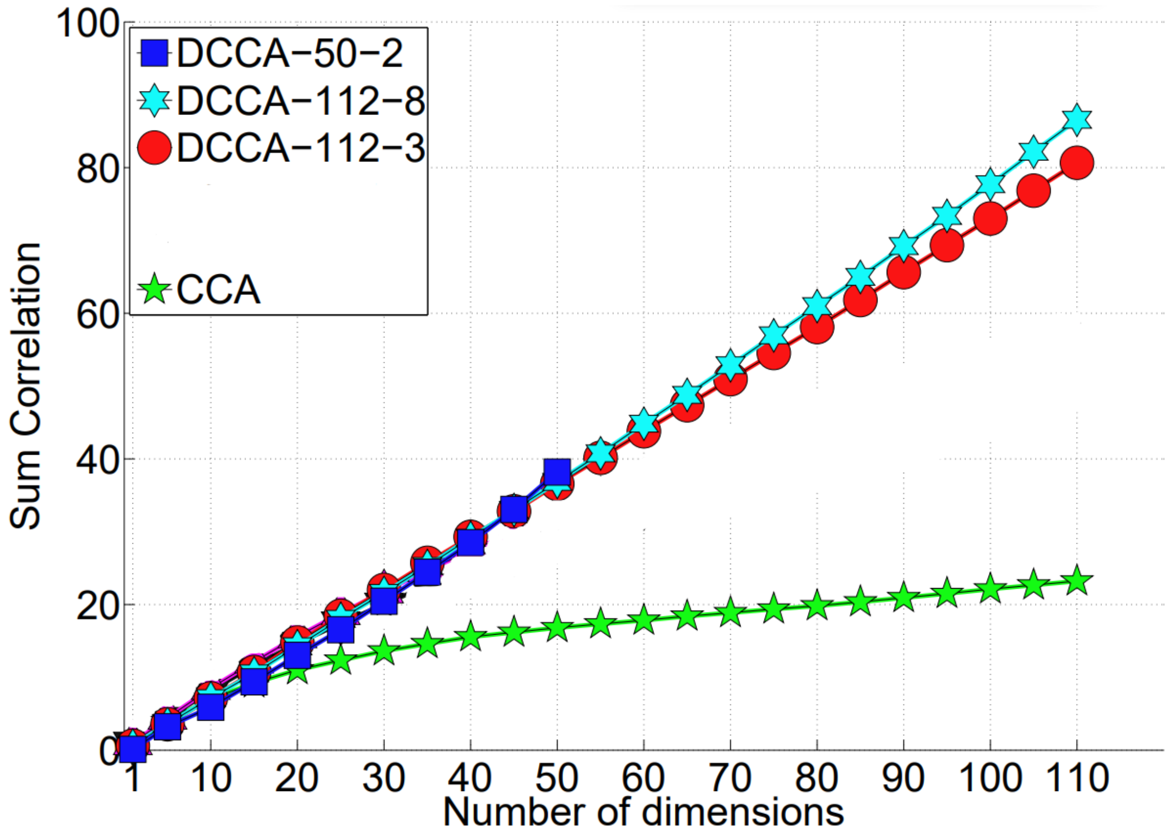
```
def svm_classify(data, C):
    """
    trains a linear SVM on the data
    input C specifies the penalty factor of SVM
    """
    train_data, _, train_label = data[0]
    valid_data, _, valid_label = data[1]
    test_data, _, test_label = data[2]

    print('training SVM...')
    clf = svm.LinearSVC(C=C, dual=False)
    clf.fit(train_data, train_label.ravel())

    p = clf.predict(test_data)
    test_acc = accuracy_score(test_label, p)
    p = clf.predict(valid_data)
    valid_acc = accuracy_score(valid_label, p)
```

```
    return [test_acc, valid_acc]
```

5. Kết quả và đánh giá:

Kết quả tổng các hệ số tương quan trên fearture output 50 chiều của DCCA và CCA trên tập dữ liệu gốc.

	CCA	DCCA (50-2)
Fold 1	16.8	**38.2**
Fold 2	15.8	**34.1**
Fold 3	16.9	**39.4**
Fold 4	16.6	**37.1**
Fold 5	16.2	**34.0**



Sử dụng độ đo ACC:

$$accuracy(y,\hat{y}) = \max_{perm \in P} \frac{1}{n} \sum_{i=0}^{n-1} 1(perm(\hat{y}_i) = y_i)$$

where P is the set of all permutations in $[1;K]$ where K is the number of clusters.

Sử dụng độ đo NMI:

- Normalized Mutual Information:

$$NMI(Y,C) = \frac{2 \times I(Y;C)}{[H(Y) + H(C)]}$$

where,

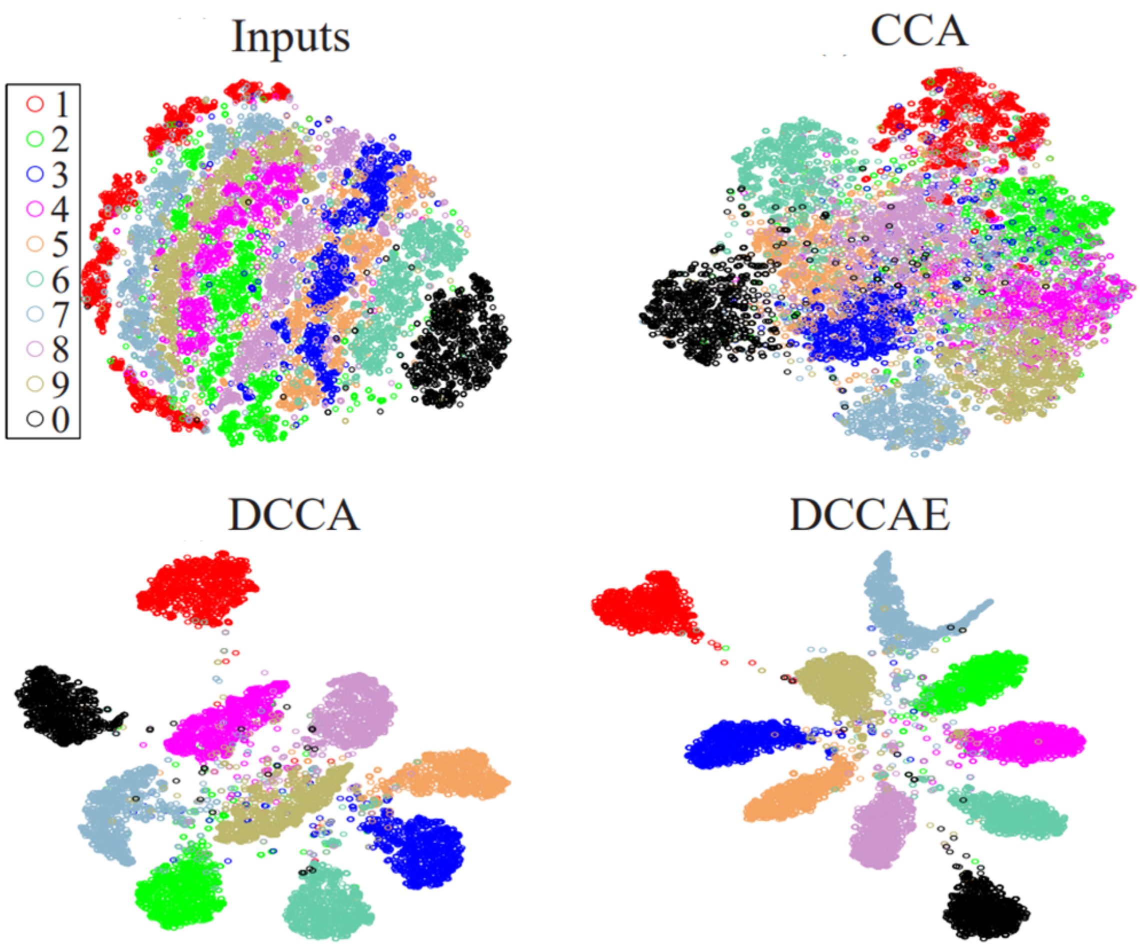
1) Y = class labels
2) C = cluster labels
3) H(.) = Entropy
4) I(Y;C) = Mutual Information b/w Y and C

Note: All logs are base-2.

Kết quả thu được:

Method	ACC (%)	NMI (%)	Error (%)
Baseline	47.0	50.6	13.1
CCA ($L = 10$)	72.9	56.0	19.6
DCCA ($L = 10$)	**97.0**	**92.0**	**2.9**
DCCAE ($L = 10$)	**97.5**	**93.4**	**2.2**

Cho thấy sau khi qua phương pháp áp dụng CCA vào mô hình tập dữ liệu input đã được tách bạch ra rõ ràng hơn. (Ở CCA: áp dụng thẳng vào dữ liệu input thì dù dữ liệu đã được phân cụm những vẫn còn quá nhiễu để có thể phân lớp, do cấu trúc của dữ liệu quá phức tạp).



Sau khi qua DCCA và DCCAE (một phiên bản mở rộng của DCCA) thì feature đã được tách bạch rõ ràng và mang lại hiệu năng cao khi phân lớp.

$$\min_{\mathbf{W_f},\mathbf{W}_g,\mathbf{W_p},\mathbf{W_q},\mathbf{U},\mathbf{V}} -\frac{1}{N}\operatorname{tr}\left(\mathbf{U}^\top \mathbf{f}(\mathbf{X})\mathbf{g}(\mathbf{Y})^\top \mathbf{V}\right)$$
$$+\frac{\lambda}{N}\sum_{i=1}^{N}\left(\|\mathbf{x}_i-\mathbf{p}(\mathbf{f}(\mathbf{x}_i))\|^2+\|\mathbf{y}_i-\mathbf{q}(\mathbf{g}(\mathbf{y}_i))\|^2\right) \quad (2)$$
$$\text{s.t.} \quad \text{the same constraints in } (1),$$

Đây là hàm mất mát của DCCAE với p(.) q(.) là hàm tái tạo output features.

6. <u>Kết quả thực nghiệm</u>

Kết quả được thực nghiệm trên 3 model khác nhau trên môi trường google colab (runtime: GPU mang lại tốc độ cao hơn CPU).

Mô hình	Accuracy			Training Time
	Train	Validation	Test	
Deep CCA	97.528%	95.8%	95.72%	90s
Multi-layer Perceptron	99.64%	100%	90.34%	50s
Linear SVM	81.01%	81.75%	81.14%	24.66s

Kết quả thực nghiệm trên mô hình Deep CCA mang lại hiệu quả tốt nhất khi so với các mô hình khác.
Với mô hình Linear SVM sử dụng trực tiếp dữ liệu thô ban đầu để thực hiện phân lớp trên bộ dữ liệu không nhiễu (View1) khiến cho tốc độ thực thi nhanh nhất nhưng mang lại kết quả không quá cao so với 2 mô hình còn lại.
Mô hình Multi-layer Perceptron sử dụng 3 hidden layer (1024, 1024, 1024) và hàm kích hoạt softmax ở layer cuối để tiến hành phân lớp cho lớp fully-connected gồm 10 lớp tương ứng với 10 chữ số viết tay mang lại độ chính xác lên đến 99.64% ở tập training, tuy nhiên chỉ đạt 90.34% trên tập test. Có thể thấy rằng dường như mô hình có sự chênh lệch cao về độ chính xác giữa tập train và tập test (overfitting).
Và cuối cùng,mô hình Deep CCA gồm 3 hidden layer (1024, 1024, 1024) và output features là vector 10 chiều để thực hiện CCA và sau đó sử dụng new features vào Linear SVM để tiến hành phân lớp. Mô hình Deep CCA cho thấy sự đồng đều về độ chính xác giữa các tập train, validation và test. Tuy chi phí traning gấp hơn 3 lần Linear SVM và gần gấp đôi Multi-layer Perceptron nhưng mang lại kết quả đáng ngưỡng mộ. Có thể thấy tiềm năng khi áp dụng CCA vào các mô hình học sâu là cực lớn.

7. Kết luận

Việc sử dụng CCA trong các mô hình học sâu mang lại những kết quả đáng mong đợi khi trích chọn các đặc trưng có độ tương quan cao giữa các view dữ liệu khác nhau trước khi đưa vào hàm phân lớp trong khi các mô hình khác thường chỉ thực hiện phân lớp trên một view dữ liệu ban đầu. Hiện nay với sự phát triển của các mô hình học sâu và dữ liệu lớn, CCA càng có vai trò quan trọng, hứa hẹn nên một sự kết hợp tốt giữa các kĩ thuật máy học và kĩ thuật thống kê.

C. Tài liệu tham khảo

1. http://proceedings.mlr.press/v28/andrew13.pdf
2. lib.uet.vnu.edu.vn/bitstream/123456789/955/1/LuanVan_HongVT.pdf
3. Michaelvll/DeepCCA: An implementation of Deep Canonical Correlation Analysis (DCCA or Deep CCA) with pytorch. (github.com)
4. Applied Multivariate Statistical Analysis 6th Edition – 2008.
5. Deep Canonical Correlation Analysis – 2013.
6. SVM with MNIST (dmkothari.github.io)
7. https://arxiv.org/pdf/1908.05349.pdf
8. Deep Probabilistic Canonical Correlation Analysis | Proceedings of the AAAI Conference on Artificial Intelligence

Hết