

Statystyczna Eksploracja Danych

Wykład 9 - analiza czynnikowa, kanoniczna analiza korelacji,
metody jądrowe w SED

dr inż. Julian Sienkiewicz

23 maja 2019

Analiza czynnikowa (*factor analysis* [FA]) - historia

- jej początki (podobnie jak analizy składowych głównych) sięgają pierwszych lat XX wieku,
- zaczęła się od pracy Charlesa Spearmana nad wykazaniem, że za naszym powodzeniem (lub jego brakiem) w każdej aktywności intelektualnej kryje się **bezpośrednio niemierzalny**, ale obiektywnie występujący **czynnik**, nazywany przez badacza ogólną inteligencją".

Analiza czynnikowa (*factor analysis* [FA]) - historia

- jej początki (podobnie jak analizy składowych głównych) sięgają pierwszych lat XX wieku,
- zaczęła się od pracy Charlesa Spearmana nad wykazaniem, że za naszym powodzeniem (lub jego brakiem) w każdej aktywności intelektualnej kryje się **bezpośrednio niemierzalny**, ale obiektywnie występujący **czynnik**, nazywany przez badacza ogólną inteligencją".

Analiza czynnikowa - cele

- znajdowanie cech (zmiennych) **ukrytych**, czyli nieobserwowanych, ale mających dobrze określony i bezpośredni wpływ na mierzone wartości **cech** obserwowanych (czyli zmiennych)

Analiza czynnikowa (*factor analysis* [FA]) - historia

- jej początki (podobnie jak analizy składowych głównych) sięgają pierwszych lat XX wieku,
- zaczęła się od pracy Charlesa Spearmana nad wykazaniem, że za naszym powodzeniem (lub jego brakiem) w każdej aktywności intelektualnej kryje się **bezpośrednio niemierzalny**, ale obiektywnie występujący **czynnik**, nazywany przez badacza ogólną inteligencją".

Analiza czynnikowa - cele

- znajdowanie cech (zmiennych) **ukrytych**, czyli nieobserwowanych, ale mających dobrze określony i bezpośredni wpływ na mierzone wartości **cech** obserwowanych (czyli zmiennych)
- w pewien sposób powiązana ze składowymi głównymi, ale w odróżnieniu od PCA stanowi **model danych**.

Wprowadźmy następujące oznaczenia:

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

$$\mathbf{x} = [x^{(1)}, \dots, x^{(p)}]^T \in \mathbb{R}^p$$

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

$$\mathbf{x} = [x^{(1)}, \dots, x^{(p)}]^T \in \mathbb{R}^p$$

Wektor obserwacji \mathbf{x} jest wektorem losowym o wektorze wartości oczekiwanych \mathbf{m} i macierzy kowariancji Σ .

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

$$\mathbf{x} = [x^{(1)}, \dots, x^{(p)}]^T \in \mathbb{R}^p$$

Wektor obserwacji \mathbf{x} jest wektorem losowym o wektorze wartości oczekiwanych \mathbf{m} i macierzy kowariancji Σ .

Wektor ukrytych czynników

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

$$\mathbf{x} = [x^{(1)}, \dots, x^{(p)}]^T \in \mathbb{R}^p$$

Wektor obserwacji \mathbf{x} jest wektorem losowym o wektorze wartości oczekiwanych \mathbf{m} i macierzy kowariancji Σ .

Wektor ukrytych czynników

$$\mathbf{z} = [z^{(1)}, \dots, z^{(k)}]^T \in \mathbb{R}^k$$

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

$$\mathbf{x} = [x^{(1)}, \dots, x^{(p)}]^T \in \mathbb{R}^p$$

Wektor obserwacji \mathbf{x} jest wektorem losowym o wektorze wartości oczekiwanych \mathbf{m} i macierzy kowariancji Σ .

Wektor ukrytych czynników

$$\mathbf{z} = [z^{(1)}, \dots, z^{(k)}]^T \in \mathbb{R}^k$$

przy czym $k < p$.

Wprowadźmy następujące oznaczenia:

Wektor obserwacji

$$\mathbf{x} = [x^{(1)}, \dots, x^{(p)}]^T \in \mathbb{R}^p$$

Wektor obserwacji \mathbf{x} jest wektorem losowym o wektorze wartości oczekiwanych \mathbf{m} i macierzy kowariancji Σ .

Wektor ukrytych czynników

$$\mathbf{z} = [z^{(1)}, \dots, z^{(k)}]^T \in \mathbb{R}^k$$

przy czym $k < p$.

Stawiamy hipotezę, że wektor obserwacji \mathbf{x} można modelować jako liniowe kombinacje mniejszego zbioru k ukrytych zmiennych losowych \mathbf{z} . Zmienne \mathbf{z} są **nieobserwowane**.

Model analizy czynnikowej

Model analizy czynnikowej

$$\mathbf{x} = \mathbf{m} + \Delta \mathbf{z} + \mathbf{e}$$

Model analizy czynnikowej

$$\mathbf{x} = \mathbf{m} + \Delta \mathbf{z} + \mathbf{e}$$

W powyższym równaniu $\Delta_{(p,k)}$ jest macierzą stałych, nieznanych współczynników δ_{ij} , $i = 1, \dots, p, j = 1, \dots, k$, natomiast \mathbf{e} jest wektorem losowym, często nazywanym **wektorem błędów**.

Model analizy czynnikowej

$$\mathbf{x} = \mathbf{m} + \Delta \mathbf{z} + \mathbf{e}$$

W powyższym równaniu $\Delta_{(p,k)}$ jest macierzą stałych, nieznanych współczynników δ_{ij} , $i = 1, \dots, p$, $j = 1, \dots, k$, natomiast \mathbf{e} jest wektorem losowym, często nazywanym **wektorem błędów**.

$$\mathbf{e} = [e^{(1)}, \dots, e^{(p)}]^T \in \mathbb{R}^p$$

Model analizy czynnikowej

$$\mathbf{x} = \mathbf{m} + \Delta \mathbf{z} + \mathbf{e}$$

W powyższym równaniu $\Delta_{(p,k)}$ jest macierzą stałych, nieznanych współczynników δ_{ij} , $i = 1, \dots, p$, $j = 1, \dots, k$, natomiast \mathbf{e} jest wektorem losowym, często nazywanym **wektorem błędów**.

$$\mathbf{e} = [e^{(1)}, \dots, e^{(p)}]^T \in \mathbb{R}^p$$

Poczynione są następujące założenia:

Model analizy czynnikowej

$$\mathbf{x} = \mathbf{m} + \Delta \mathbf{z} + \mathbf{e}$$

W powyższym równaniu $\Delta_{(p,k)}$ jest macierzą stałych, nieznanych współczynników δ_{ij} , $i = 1, \dots, p, j = 1, \dots, k$, natomiast \mathbf{e} jest wektorem losowym, często nazywanym **wektorem błędów**.

$$\mathbf{e} = [e^{(1)}, \dots, e^{(p)}]^T \in \mathbb{R}^p$$

Poczynione są następujące założenia:

$$\begin{array}{lll} E(\mathbf{z}) = \mathbf{0} & \text{cov}(\mathbf{z}) = \mathbf{I} \\ E(\mathbf{e}) = \mathbf{0} & \text{cov}(\mathbf{e}) = \Psi & \text{cov}(\mathbf{z}, \mathbf{e}) = \mathbf{0} \end{array}$$

Model analizy czynnikowej

$$\mathbf{x} = \mathbf{m} + \Delta \mathbf{z} + \mathbf{e}$$

W powyższym równaniu $\Delta_{(p,k)}$ jest macierzą stałych, nieznanych współczynników δ_{ij} , $i = 1, \dots, p, j = 1, \dots, k$, natomiast \mathbf{e} jest wektorem losowym, często nazywanym **wektorem błędów**.

$$\mathbf{e} = [e^{(1)}, \dots, e^{(p)}]^T \in \mathbb{R}^p$$

Poczynione są następujące założenia:

$$\begin{aligned} E(\mathbf{z}) &= \mathbf{0} & \text{cov}(\mathbf{z}) &= \mathbf{I} \\ E(\mathbf{e}) &= \mathbf{0} & \text{cov}(\mathbf{e}) &= \Psi & \text{cov}(\mathbf{z}, \mathbf{e}) &= \mathbf{0} \end{aligned}$$

gdzie Ψ jest macierzą diagonalną o elementach Ψ_{jj} na przekątnej.

Definicje

Każdą współrzędną scentrowanego wektora obserwacji można przedstawić jako sumę kombinacji liniowej k czynników oraz nieskorelowanej z czynnikami zmiennej losowej $e^{(i)}$

Każdą współrzędną scentrowanego wektora obserwacji można przedstawić jako sumę kombinacji liniowej k czynników oraz nieskorelowanej z czynnikami zmiennej losowej $e^{(i)}$

$$x^{(i)} - m^{(i)} = \sum_{j=1}^k \delta_{ij} z^{(j)} + e^{(i)}$$

Definicje

Każdą współrzędną scentrowanego wektora obserwacji można przedstawić jako sumę kombinacji liniowej k czynników oraz nieskorelowanej z czynnikami zmiennej losowej $e^{(i)}$

$$x^{(i)} - m^{(i)} = \sum_{j=1}^k \delta_{ij} z^{(j)} + e^{(i)}$$

Terminologia

Zmienne $e^{(i)}$ nazywamy **czynnikami specyficznymi**.

Czynniki $z^{(i)}$ noszą nazwę **czynników wspólnych**.

Jest to związane z faktem, iż i -ty czynnik specyficzny ma wpływ jedynie na i -tą składową wektora obserwacji, podczas gdy czynniki wspólne wyznaczają korelacje, istniejące pomiędzy różnymi zmiennymi.

Współczynniki δ_{ij} noszą nazwę **ładunków czynników** - wektor $[\delta_{1j}, \dots, \delta_{pj}]^T$ to wektor ładunków j -tego czynnika.

Definicje

Każdą współrzędną scentrowanego wektora obserwacji można przedstawić jako sumę kombinacji liniowej k czynników oraz nieskorelowanej z czynnikami zmiennej losowej $e^{(i)}$

$$x^{(i)} - m^{(i)} = \sum_{j=1}^k \delta_{ij} z^{(j)} + e^{(i)}$$

Terminologia

Zmienne $e^{(i)}$ nazywamy **czynnikami specyficznymi**.

Czynniki $z^{(i)}$ noszą nazwę **czynników wspólnych**.

Jest to związane z faktem, iż i -ty czynnik specyficzny ma wpływ jedynie na i -tą składową wektora obserwacji, podczas gdy czynniki wspólne wyznaczają korelacje, istniejące pomiędzy różnymi zmiennymi.

Współczynniki δ_{ij} noszą nazwę **ładunków czynników** - wektor $[\delta_{1j}, \dots, \delta_{pj}]^T$ to wektor ładunków j -tego czynnika.

Model analizy czynnikowej można zapisać schematycznie jako

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane =

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

Model analizy czynnikowej można zapisać schematycznie jako

$$\begin{array}{|c|} \hline \text{zmienne} \\ \hline \text{obserwowane} \\ \hline \end{array} = \begin{array}{|c|} \hline \text{ładunki} \\ \hline \end{array} \times$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$\Sigma =$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\Sigma = \text{cov}(\mathbf{x}) =$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\Sigma = \text{cov}(\mathbf{x}) = \text{cov}(\Delta \mathbf{z} + \mathbf{e}) =$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\Sigma = \text{cov}(\mathbf{x}) = \text{cov}(\Delta \mathbf{z} + \mathbf{e}) = \text{cov}(\Delta \mathbf{z}) + \text{cov}(\mathbf{e}) =$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\begin{aligned}\Sigma &= \text{cov}(\mathbf{x}) = \text{cov}(\Delta \mathbf{z} + \mathbf{e}) = \text{cov}(\Delta \mathbf{z}) + \text{cov}(\mathbf{e}) = \\ &= \Delta \text{cov}(\mathbf{z}) \text{cov}(\mathbf{z}^T) \Delta^T + \text{cov}(\mathbf{e})\end{aligned}$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\begin{aligned}\Sigma &= \text{cov}(\mathbf{x}) = \text{cov}(\Delta \mathbf{z} + \mathbf{e}) = \text{cov}(\Delta \mathbf{z}) + \text{cov}(\mathbf{e}) = \\ &= \Delta \text{cov}(\mathbf{z}) \text{cov}(\mathbf{z}^T) \Delta^T + \text{cov}(\mathbf{e}) = \Delta \Delta^T + \Psi\end{aligned}$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\begin{aligned}\Sigma = \text{cov}(\mathbf{x}) &= \text{cov}(\Delta \mathbf{z} + \mathbf{e}) = \text{cov}(\Delta \mathbf{z}) + \text{cov}(\mathbf{e}) = \\ &= \Delta \text{cov}(\mathbf{z}) \text{cov}(\mathbf{z}^T) \Delta^T + \text{cov}(\mathbf{e}) = \Delta \Delta^T + \Psi\end{aligned}$$

W szczególności wariancje zmiennych $x^{(i)}$ mają postać

$$\sigma_{ii} = \sum_{j=1}^k \delta_{ij}^2 + \psi_{ii}$$

Model analizy czynnikowej można zapisać schematycznie jako

zmienne
obserwowane

=

ładunki

x

czynniki
wspólne

+

czynniki
specyficzne

Policzmy teraz macierz kowariancji wektora obserwacji

$$\begin{aligned}\Sigma = \text{cov}(\mathbf{x}) &= \text{cov}(\Delta \mathbf{z} + \mathbf{e}) = \text{cov}(\Delta \mathbf{z}) + \text{cov}(\mathbf{e}) = \\ &= \Delta \text{cov}(\mathbf{z}) \text{cov}(\mathbf{z}^T) \Delta^T + \text{cov}(\mathbf{e}) = \Delta \Delta^T + \Psi\end{aligned}$$

W szczególności wariancje zmiennych $x^{(i)}$ mają postać

$$\sigma_{ii} = \sum_{j=1}^k \delta_{ij}^2 + \psi_{ii}$$

czyli są sumą **wariancji specyficznej** ψ_{ii} (związanej tylko ze zmienną $x^{(i)}$) oraz **wariancji wspólnej**.

Analiza czynnikowa sprowadza się do sprawdzenia, czy macierz kowariancji obserwowanych danych daje się adekwatnie przedstawić za pomocą rozkładu $\Sigma = \Delta\Delta^T + \Psi$, przy założeniu modelu $\mathbf{x} = \Delta\mathbf{z} + \mathbf{e}$.

Analiza czynnikowa sprowadza się do sprawdzenia, czy macierz kowariancji obserwowanych danych daje się adekwatnie przedstawić za pomocą rozkładu $\Sigma = \Delta\Delta^T + \Psi$, przy założeniu modelu $\mathbf{x} = \Delta\mathbf{z} + \mathbf{e}$.

Otrzymanie odpowiedzi twierdzącej dla pewnej liczby k czynników wspólnych pozwala stwierdzić, że wektor p obserwacji jest w efekcie *rządzony* przez k ukrytych czynników, których znaczenie musi być oddzielnie określone. Interpretacja polega na wykryciu tego, co ukryte czynniki mogą oznaczać.

Analiza czynnikowa sprowadza się do sprawdzenia, czy macierz kowariancji obserwowanych danych daje się adekwatnie przedstawić za pomocą rozkładu $\Sigma = \Delta\Delta^T + \Psi$, przy założeniu modelu $\mathbf{x} = \Delta\mathbf{z} + \mathbf{e}$.

Otrzymanie odpowiedzi twierdzącej dla pewnej liczby k czynników wspólnych pozwala stwierdzić, że wektor p obserwacji jest w efekcie *rządzony* przez k ukrytych czynników, których znaczenie musi być oddzielnie określone. Interpretacja polega na wykryciu tego, co ukryte czynniki mogą oznaczać.

Ze strony technicznej problem sprowadza się do kwestii estymacji nieznanych parametrów δ_{ij} oraz ψ_{ij} .

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie.

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie. Pomnożenie ich wektora przez dowolną ustaloną macierz ortogonalną nie zmienia postaci modelu, gdyż

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie. Pomnożenie ich wektora przez dowolną ustaloną macierz ortogonalną nie zmienia postaci modelu, gdyż

$$\mathbf{x} = \mathbf{m} + (\Delta \mathbf{G})(\mathbf{G}^T \mathbf{z})$$

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie. Pomnożenie ich wektora przez dowolną ustaloną macierz ortogonalną nie zmienia postaci modelu, gdyż

$$\mathbf{x} = \mathbf{m} + (\Delta \mathbf{G})(\mathbf{G}^T \mathbf{z})$$

gdzie $\mathbf{G}_{(k,k)}$ jest dowolną macierzą ortogonalną.

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie. Pomnożenie ich wektora przez dowolną ustaloną macierz ortogonalną nie zmienia postaci modelu, gdyż

$$\mathbf{x} = \mathbf{m} + (\Delta \mathbf{G})(\mathbf{G}^T \mathbf{z})$$

gdzie $\mathbf{G}_{(k,k)}$ jest dowolną macierzą ortogonalną.

W efekcie zmiana macierzy Δ na $\Delta \mathbf{G}$ nie prowadzi do zmiany postaci kowariancji Σ :

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie. Pomnożenie ich wektora przez dowolną ustaloną macierz ortogonalną nie zmienia postaci modelu, gdyż

$$\mathbf{x} = \mathbf{m} + (\Delta \mathbf{G})(\mathbf{G}^T \mathbf{z})$$

gdzie $\mathbf{G}_{(k,k)}$ jest dowolną macierzą ortogonalną.

W efekcie zmiana macierzy Δ na $\Delta \mathbf{G}$ nie prowadzi do zmiany postaci kowariancji Σ :

$$\Sigma = (\Delta \mathbf{G})(\Delta \mathbf{G})^T + \Psi = \Delta \Delta^T + \Psi$$

Należy od razu zauważyć, iż czynniki ukryte w zaproponowanym modelu nie mogą zostać wyznaczone jednoznacznie. Pomnożenie ich wektora przez dowolną ustaloną macierz ortogonalną nie zmienia postaci modelu, gdyż

$$\mathbf{x} = \mathbf{m} + (\Delta \mathbf{G})(\mathbf{G}^T \mathbf{z})$$

gdzie $\mathbf{G}_{(k,k)}$ jest dowolną macierzą ortogonalną.

W efekcie zmiana macierzy Δ na $\Delta \mathbf{G}$ nie prowadzi do zmiany postaci kowariancji Σ :

$$\Sigma = (\Delta \mathbf{G})(\Delta \mathbf{G})^T + \Psi = \Delta \Delta^T + \Psi$$

Takie pomnożenie lewostronnie przez macierz \mathbf{G} odpowiada **obrotowi** czynników w przestrzeni \mathbb{R}^k . Prowadzi to do odpowiedniej zmiany wartości ładunków, ale żadnej zmianie nie podlega rozkład macierzy Σ

Innymi słowy model $\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$ jest niedookreślony, gdyż macierz Δ nie daje się jednoznacznie wyznaczyć.

Innymi słowy model $\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$ jest niedookreślony, gdyż macierz Δ nie daje się jednoznacznie wyznaczyć.

Powszechnie przyjętym rozwiązaniem jest narzucenie dodatkowych warunków na elementy macierzy Δ lub macierzy Δ i Ψ tak, aby otrzymać jednoznaczność. Jednym z najprostszych (całkowicie arbitralnym) warunkiem jest zażądanie, aby macierz $\Delta \Delta^T$ była diagonalna.

Innymi słowy model $\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$ jest niedookreślony, gdyż macierz Δ nie daje się jednoznacznie wyznaczyć.

Powszechnie przyjętym rozwiązaniem jest narzucenie dodatkowych warunków na elementy macierzy Δ lub macierzy Δ i Ψ tak, aby otrzymać jednoznaczność. Jednym z najprostszych (całkowicie arbitralnym) warunkiem jest zażądanie, aby macierz $\Delta \Delta^T$ była diagonalna.

Otrzymawszy jako macierz ładunków Δ i macierz Ψ można swobodnie dokonać dowolnego obrotu wektora czynników ukrytych — jest to nam na rękę, gdyż pozwala na wybranie takiego obrotu, by nowy wektor $\mathbf{G}^T \mathbf{z}$ dawał ładunki umożliwiające **jak najlepszą interpretację** nowych czynników.

Oszacujmy teraz górne ograniczenie na liczbę czynników wspólnych. W równaniu na Σ mamy

$$p(p + 1)/2$$

Oszacujmy teraz górne ograniczenie na liczbę czynników wspólnych. W równaniu na Σ mamy

$$p(p+1)/2$$

znanych parametrów macierzy Σ oraz

$$p + pk$$

Oszacujemy teraz górne ograniczenie na liczbę czynników wspólnych. W równaniu na Σ mamy

$$p(p+1)/2$$

znanych parametrów macierzy Σ oraz

$$p + pk$$

parametrów **nieznanych** (p wariancji specyficznych i pk ładunków δ_{ij}).

Estymacja

Oszacujmy teraz górne ograniczenie na liczbę czynników wspólnych. W równaniu na Σ mamy

$$p(p + 1)/2$$

znanych parametrów macierzy Σ oraz

$$p + pk$$

parametrów **nieznanych** (p wariancji specyficznych i pk ładunków δ_{ij}). Ponadto diagonalizacja macierzy $\Delta\Delta^T$ daje

$$k(k - 1)/2$$

Estymacja

Oszacujmy teraz górne ograniczenie na liczbę czynników wspólnych. W równaniu na Σ mamy

$$p(p + 1)/2$$

znanych parametrów macierzy Σ oraz

$$p + pk$$

parametrów **nieznanych** (p wariancji specyficznych i pk ładunków δ_{ij}). Ponadto diagonalizacja macierzy $\Delta\Delta^T$ daje

$$k(k - 1)/2$$

dodatkowych więzów.

Oszacujemy teraz górne ograniczenie na liczbę czynników wspólnych. W równaniu na Σ mamy

$$p(p + 1)/2$$

znanych parametrów macierzy Σ oraz

$$p + pk$$

parametrów **nieznanych** (p wariacji specyficznych i pk ładunków δ_{ij}). Ponadto diagonalizacja macierzy $\Delta\Delta^T$ daje

$$k(k - 1)/2$$

dodatkowych więzów. Ostatecznie zachodzi nierówność

$$p(p + 1)/2 \geq p(k + 1) - k(k - 1)/2$$

W efekcie liczba k czynników wspólnych musi spełniać nierówność

$$(p - k)^2 \geq p + k$$

W efekcie liczba k czynników wspólnych musi spełniać nierówność

$$(p - k)^2 \geq p + k$$

Oczywiście w praktyce chcemy, aby liczba czynników ukrytych odpowiadających za strukturę danych była mała, co oznacza, że zawsze spełniamy powyższą nierówność z dużym zapasem.

W efekcie liczba k czynników wspólnych musi spełniać nierówność

$$(p - k)^2 \geq p + k$$

Oczywiście w praktyce chcemy, aby liczba czynników ukrytych odpowiadających za strukturę danych była mała, co oznacza, że zawsze spełniamy powyższą nierówność z dużym zapasem. W celu estymacji elementów macierzy Δ i Ψ posługujemy się metodą najmniejszych kwadratów. W tym celu minimalizujemy sumę

W efekcie liczba k czynników wspólnych musi spełniać nierówność

$$(p - k)^2 \geq p + k$$

Oczywiście w praktyce chcemy, aby liczba czynników ukrytych odpowiadających za strukturę danych była mała, co oznacza, że zawsze spełniamy powyższą nierówność z dużym zapasem.

W celu estymacji elementów macierzy Δ i Ψ posługujemy się metodą najmniejszych kwadratów. W tym celu minimalizujemy sumę

$$\sum_{i=1}^p \sum_{j=1}^p (s_{ij} - \sigma_{ij})^2$$

W efekcie liczba k czynników wspólnych musi spełniać nierówność

$$(p - k)^2 \geq p + k$$

Oczywiście w praktyce chcemy, aby liczba czynników ukrytych odpowiadających za strukturę danych była mała, co oznacza, że zawsze spełniamy powyższą nierówność z dużym zapasem.

W celu estymacji elementów macierzy Δ i Ψ posługujemy się metodą najmniejszych kwadratów. W tym celu minimalizujemy sumę

$$\sum_{i=1}^p \sum_{j=1}^p (s_{ij} - \sigma_{ij})^2$$

gdzie s_{ij} jest elementem macierzy \mathbf{S} a σ_{ij} macierzy Σ .

Algorytm estymacji

Algorytm estymacji

1. zaczynamy od policzenia pochodnych cząstkowych z kryterium i przyrównania do zera,

Algorytm estymacji

- 1 zaczynamy od policzenia pochodnych cząstkowych z kryterium i przyrównania do zera,
- 2 obliczamy pierwsze przybliżenie macierzy Ψ , które oznaczamy jako $\hat{\Psi}_0$,

Algorytm estymacji

- 1 zaczynamy od policzenia pochodnych cząstkowych z kryterium i przyrównania do zera,
- 2 obliczamy pierwsze przybliżenie macierzy Ψ , które oznaczamy jako $\hat{\Psi}_0$,
- 3 jako pierwsze przybliżenie macierzy Δ bierzemy macierz k wektorów własnych macierzy $\mathbf{S} - \hat{\Psi}_0$, odpowiadających k największym wartościom własnym tej macierzy,

Algorytm estymacji

- 1 zaczynamy od policzenia pochodnych cząstkowych z kryterium i przyrównania do zera,
- 2 obliczamy pierwsze przybliżenie macierzy Ψ , które oznaczamy jako $\hat{\Psi}_0$,
- 3 jako pierwsze przybliżenie macierzy Δ bierzemy macierz k wektorów własnych macierzy $\mathbf{S} - \hat{\Psi}_0$, odpowiadających k największym wartościom własnym tej macierzy,
- 4 procedurę kontynuujemy aż do uzyskania zbieżności algorytmu (kolejne przybliżenia macierzy Ψ są równe różnicy macierzy \mathbf{S} i aktualnego przybliżenia macierzy Δ).

Równania określające model analizy czynnikowej oraz składowe główne są dość podobne. Faktycznie, jeżeli założyć, że dane są scentrowane, mamy

Równania określające model analizy czynnikowej oraz składowe główne są dość podobne. Faktycznie, jeżeli założyć, że dane są scentrowane, mamy

$$\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$$

Równania określające model analizy czynnikowej oraz składowe główne są dość podobne. Faktycznie, jeżeli założyć, że dane są scentrowane, mamy

$$\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$$

oraz

$$\mathbf{y} = \Gamma^T \mathbf{x}$$

Równania określające model analizy czynnikowej oraz składowe główne są dość podobne. Faktycznie, jeżeli założyć, że dane są scentrowane, mamy

$$\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$$

oraz

$$\mathbf{y} = \Gamma^T \mathbf{x}$$

Ponieważ Γ jest ortogonalna, to

$$\mathbf{x} = \Gamma \mathbf{y}$$

Równania określające model analizy czynnikowej oraz składowe główne są dość podobne. Faktycznie, jeżeli założyć, że dane są scentrowane, mamy

$$\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$$

oraz

$$\mathbf{y} = \Gamma^T \mathbf{x}$$

Ponieważ Γ jest ortogonalna, to

$$\mathbf{x} = \Gamma \mathbf{y}$$

co można rozpisać jako

Równania określające model analizy czynnikowej oraz składowe główne są dość podobne. Faktycznie, jeżeli założyć, że dane są scentrowane, mamy

$$\mathbf{x} = \Delta \mathbf{z} + \mathbf{e}$$

oraz

$$\mathbf{y} = \Gamma^T \mathbf{x}$$

Ponieważ Γ jest ortogonalna, to

$$\mathbf{x} = \Gamma \mathbf{y}$$

co można rozpisać jako

$$x^{(1)} = \gamma_{11}y^{(1)} + \dots + \gamma_{1p}y^{(p)}$$

$$\vdots$$

$$x^{(p)} = \gamma_{p1}y^{(1)} + \dots + \gamma_{pp}y^{(p)}$$

Potraktowanie ostatnich $p - k$ składników prawych stron jako szumu $e^{(i)}$ pozwala je zapisać jako

Potraktowanie ostatnich $p - k$ składników prawych stron jako szumu $e^{(i)}$ pozwala je zapisać jako

$$x^{(1)} = \gamma_{11}y^{(1)} + \dots + \gamma_{1k}y^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \gamma_{p1}y^{(1)} + \dots + \gamma_{pk}y^{(k)} + e^{(p)}$$

Potraktowanie ostatnich $p - k$ składników prawych stron jako szumu $e^{(i)}$ pozwala je zapisać jako

$$x^{(1)} = \gamma_{11}y^{(1)} + \dots + \gamma_{1k}y^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \gamma_{p1}y^{(1)} + \dots + \gamma_{pk}y^{(k)} + e^{(p)}$$

lub

Potraktowanie ostatnich $p - k$ składników prawych stron jako szumu $e^{(i)}$ pozwala je zapisać jako

$$x^{(1)} = \gamma_{11}y^{(1)} + \dots + \gamma_{1k}y^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \gamma_{p1}y^{(1)} + \dots + \gamma_{pk}y^{(k)} + e^{(p)}$$

lub

$$x^{(1)} = \delta_{11}z^{(1)} + \dots + \delta_{1k}z^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \delta_{p1}y^{(1)} + \dots + \delta_{pk}z^{(k)} + e^{(p)}$$

Potraktowanie ostatnich $p - k$ składników prawych stron jako szumu $e^{(i)}$ pozwala je zapisać jako

$$x^{(1)} = \gamma_{11}y^{(1)} + \dots + \gamma_{1k}y^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \gamma_{p1}y^{(1)} + \dots + \gamma_{pk}y^{(k)} + e^{(p)}$$

lub

$$x^{(1)} = \delta_{11}z^{(1)} + \dots + \delta_{1k}z^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \delta_{p1}y^{(1)} + \dots + \delta_{pk}z^{(k)} + e^{(p)}$$

gdzie $\delta_{ij} = \gamma_{ij}\sqrt{\lambda_j}$ oraz $z^{(j)} = y^{(j)}/\sqrt{\lambda_j}$ i λ_j są wartościami własnymi macierzy $\Sigma = \Gamma\Lambda\Gamma^T$.

Potraktowanie ostatnich $p - k$ składników prawych stron jako szumu $e^{(i)}$ pozwala je zapisać jako

$$x^{(1)} = \gamma_{11}y^{(1)} + \dots + \gamma_{1k}y^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \gamma_{p1}y^{(1)} + \dots + \gamma_{pk}y^{(k)} + e^{(p)}$$

lub

$$x^{(1)} = \delta_{11}z^{(1)} + \dots + \delta_{1k}z^{(k)} + e^{(1)}$$

$$\vdots$$

$$x^{(p)} = \delta_{p1}y^{(1)} + \dots + \delta_{pk}z^{(k)} + e^{(p)}$$

gdzie $\delta_{ij} = \gamma_{ij}\sqrt{\lambda_j}$ oraz $z^{(j)} = y^{(j)}/\sqrt{\lambda_j}$ i λ_j są wartościami własnymi macierzy $\Sigma = \Gamma\Lambda\Gamma^T$. Formalnie mamy więc taką samą postać równań.

Czy w takim razie zadanie analizy czynnikowej jest równoważne
zadaniu znalezienia składowych głównych?

Czy w takim razie zadanie analizy czynnikowej jest równoważne
zadaniu znalezienia składowych głównych?

NIE!

Czy w takim razie zadanie analizy czynnikowej jest równoważne zadaniu znalezienia składowych głównych?

NIE! Zmienne $e^{(j)}$ są nieskorelowane.

Czy w takim razie zadanie analizy czynnikowej jest równoważne zadaniu znalezienia składowych głównych?

NIE! Zmienne $e^{(j)}$ są nieskorelowane. Wyjątkiem jest sytuacja gdy $e^{(j)}$ są zerowe (czyli macierz Ψ jest równa zeru) — wtedy zadania FA i PCA są równoważne.

Czy w takim razie zadanie analizy czynnikowej jest równoważne zadaniu znalezienia składowych głównych?

NIE! Zmienne $e^{(j)}$ są nieskorelowane. Wyjątkiem jest sytuacja gdy $e^{(j)}$ są zerowe (czyli macierz Ψ jest równa zeru) — wtedy zadania FA i PCA są równoważne.

4.3 FA: Example

120 students have each taken **five exams**, the first two covering **mathematics**, the next two on **literature**, and a **comprehensive** fifth exam. It seems reasonable that the five grades for a given student ought to be related. Some students are good at both subjects, some are good at only one, etc. The goal of this analysis is to determine if there is quantitative evidence that **the students' grades on the five different exams are largely determined by only two types of ability**.

Loadings =

0.6289	0.3485
0.6992	0.3287
0.7785	-0.2069
0.7246	-0.2070
0.8963	-0.0473

A factor distinguishing
those good in mathematics
and bad in literature or
viceversa

↑
An overall factor affecting to all exams (mainly the comprehensive, then literature and finally mathematics). This can be interpreted as a general intelligence factor.

NoiseVar =

0.4829
0.4031
0.3512
0.4321
0.1944

The comprehensive
exam is the easiest to
predict with these two
factors

↑
Normalized to 1 (1=no variance reduction obtained by commonalities; 0=all variance is explained by commonalities)

4.3 FA: Example

Example:



count =

11	11	9	Traffic count in three different places (thousands/day)
7	13	11	
14	17	20	
11	13	9	
43	51	69	

...

Loadings = NoiseVar =

0.9683	0.0624
0.9636	0.0714
0.9913	0.0172

As expected X3 is highly related to the main factor, and it mostly explains the traffic in X1 and X2. The noise is coming from the red arrows.

Kanoniczna analiza korelacji (*canonical correlation analysis* — CCA) datuje się na połowe lat 30-tych. W wielu przypadkach interesują nas **liniowe korelacje** pomiędzy zbiorami zmiennych.

Kanoniczna analiza korelacji (*canonical correlation analysis* — CCA) datuje się na połowe lat 30-tych. W wielu przypadkach interesują nas **liniowe korelacje** pomiędzy zbiorami zmiennych. Niech

$$\mathbf{z} = (\mathbf{x}_i, \mathbf{y}_i), \quad \text{dla } i = 1, \dots, N$$

Kanoniczna analiza korelacji (*canonical correlation analysis* — CCA) datuje się na połowę lat 30-tych. W wielu przypadkach interesują nas **liniowe korelacje** pomiędzy zbiorami zmiennych. Niech

$$\mathbf{z} = (\mathbf{x}_i, \mathbf{y}_i), \quad \text{dla } i = 1, \dots, N$$

oznacza pomiary N obiektów, gdzie \mathbf{x}_i oraz \mathbf{y}_i mogą oznaczać pewne cechy tych obiektów.

Kanoniczna analiza korelacji (*canonical correlation analysis* — CCA) datuje się na połowę lat 30-tych. W wielu przypadkach interesują nas **liniowe korelacje** pomiędzy zbiorami zmiennych. Niech

$$\mathbf{z} = (\mathbf{x}_i, \mathbf{y}_i), \quad \text{dla } i = 1, \dots, N$$

oznacza pomiary N obiektów, gdzie \mathbf{x}_i oraz \mathbf{y}_i mogą oznaczać pewne cechy tych obiektów.

Można wyobrazić sobie np., że mamy do czynienia z n_x pomiarami zdolności czytania (zawartymi w \mathbf{x}_i) oraz n_y pomiarami zdolności analitycznych (zawartymi w \mathbf{y}_i) dla N osobników.

Kanoniczna analiza korelacji (*canonical correlation analysis* — CCA) datuje się na połowe lat 30-tych. W wielu przypadkach interesują nas **liniowe korelacje** pomiędzy zbiorami zmiennych. Niech

$$\mathbf{z} = (\mathbf{x}_i, \mathbf{y}_i), \quad \text{dla } i = 1, \dots, N$$

oznacza pomiary N obiektów, gdzie \mathbf{x}_i oraz \mathbf{y}_i mogą oznaczać pewne cechy tych obiektów.

Można wyobrazić sobie np., że mamy do czynienia z n_x pomiarami zdolności czytania (zawartymi w \mathbf{x}_i) oraz n_y pomiarami zdolności analitycznych (zawartymi w \mathbf{y}_i) dla N osobników. W zwarty sposób:

$$\mathbf{Z} = (\mathbf{X} \quad \mathbf{Y}) = \begin{pmatrix} x_{11} & \dots & x_{1n_x} & y_{11} & \dots & y_{1n_y} \\ x_{21} & \dots & x_{2n_x} & y_{21} & \dots & y_{2n_y} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{N1} & \dots & x_{Nn_x} & y_{N1} & \dots & y_{Nn_y} \end{pmatrix}$$

Kanoniczna analiza korelacji (*canonical correlation analysis* — CCA) datuje się na połowe lat 30-tych. W wielu przypadkach interesują nas **liniowe korelacje** pomiędzy zbiorami zmiennych. Niech

$$\mathbf{z} = (\mathbf{x}_i, \mathbf{y}_i), \quad \text{dla } i = 1, \dots, N$$

oznacza pomiary N obiektów, gdzie \mathbf{x}_i oraz \mathbf{y}_i mogą oznaczać pewne cechy tych obiektów.

Można wyobrazić sobie np., że mamy do czynienia z n_x pomiarami zdolności czytania (zawartymi w \mathbf{x}_i) oraz n_y pomiarami zdolności analitycznych (zawartymi w \mathbf{y}_i) dla N osobników. W zwarty sposób:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1n_x} & y_{11} & \dots & y_{1n_y} \\ x_{21} & \dots & x_{2n_x} & y_{21} & \dots & y_{2n_y} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{N1} & \dots & x_{Nn_x} & y_{N1} & \dots & y_{Nn_y} \end{pmatrix}$$

Innym dobrym przykładem mogłyby być dwa korpusy językowe (np. jeden w języku angielskim, drugi — w niemieckim). W takim wypadku elementy macierzy \mathbf{X} i \mathbf{Y} oznaczałyby po prostu ile razy dane słowo wystąpiło w konkretnym języku.

Innym dobrym przykładem mogłyby być dwa korpusy językowe (np. jeden w języku angielskim, drugi — w niemieckim). W takim wypadku elementy macierzy X i Y oznaczałyby po prostu ile razy dane słowo wystąpiło w konkretnym języku.

W obu przypadkach interesuje nas następujące zagadnienie:

Innym dobrym przykładem mogłyby być dwa korpusy językowe (np. jeden w języku angielskim, drugi — w niemieckim). W takim wypadku elementy macierzy \mathbf{X} i \mathbf{Y} oznaczałyby po prostu ile razy dane słowo wystąpiło w konkretnym języku.

W obu przypadkach interesuje nas następujące zagadnienie:

Znaleźć takie wektory \mathbf{v} i \mathbf{w} , że korelacja pomiędzy $\mathbf{a} = \mathbf{Xv}$ i $\mathbf{b} = \mathbf{Yw}$ jest maksymalna

Innym dobrym przykładem mogłyby być dwa korpusy językowe (np. jeden w języku angielskim, drugi — w niemieckim). W takim wypadku elementy macierzy \mathbf{X} i \mathbf{Y} oznaczałyby po prostu ile razy dane słowo wystąpiło w konkretnym języku.

W obu przypadkach interesuje nas następujące zagadnienie:

Znaleźć takie wektory \mathbf{v} i \mathbf{w} , że korelacja pomiędzy $\mathbf{a} = \mathbf{X}\mathbf{v}$ i $\mathbf{b} = \mathbf{Y}\mathbf{w}$ jest maksymalna

Przy założeniu, że korzystamy ze współczynnika korelacji ρ sprowadza się to do następującego sformułowania

$$\max_{\mathbf{a}, \mathbf{b}} \rho = E(\mathbf{ab}) \quad \text{przy warunkach} \quad E(\mathbf{a}^2) = 1 \quad E(\mathbf{b}^2) = 1$$

Innym dobrym przykładem mogłyby być dwa korpusy językowe (np. jeden w języku angielskim, drugi — w niemieckim). W takim wypadku elementy macierzy \mathbf{X} i \mathbf{Y} oznaczałyby po prostu ile razy dane słowo wystąpiło w konkretnym języku.

W obu przypadkach interesuje nas następujące zagadnienie:

Znaleźć takie wektory \mathbf{v} i \mathbf{w} , że korelacja pomiędzy $\mathbf{a} = \mathbf{Xv}$ i $\mathbf{b} = \mathbf{Yw}$ jest maksymalna

Przy założeniu, że korzystamy ze współczynnika korelacji ρ sprowadza się to do następującego sformułowania

$$\max_{\mathbf{a}, \mathbf{b}} \rho = E(\mathbf{ab}) \quad \text{przy warunkach} \quad E(\mathbf{a}^2) = 1 \quad E(\mathbf{b}^2) = 1$$

Warunki narzucają po prostu jednostkowe odchylenie standardowe, ponadto milcząco zakładamy, że mamy do czynienia z wyśrodkowanymi danymi (tzn. $E(\mathbf{X}) = E(\mathbf{Y}) = 0$).

Taki zapis można również przedstawić w sformułowaniu lagranżianu:

Taki zapis można również przedstawić w sformułowaniu lagranżianu:

$$\min_{\mathbf{a}, \mathbf{b}} \max_{\lambda_1, \lambda_2} \sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{y}_i \mathbf{w} - \frac{\lambda_1}{2} \left(\sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{x}_i \mathbf{v} - N \right) - \frac{\lambda_2}{2} \left(\sum_i \mathbf{w}^T \mathbf{y}_i^T \mathbf{y}_i \mathbf{w} - N \right)$$

Taki zapis można również przedstawić w sformułowaniu lagranżianu:

$$\min_{\mathbf{a}, \mathbf{b}} \max_{\lambda_1, \lambda_2} \sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{y}_i \mathbf{w} - \frac{\lambda_1}{2} \left(\sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{x}_i \mathbf{v} - N \right) - \frac{\lambda_2}{2} \left(\sum_i \mathbf{w}^T \mathbf{y}_i^T \mathbf{y}_i \mathbf{w} - N \right)$$

Oczywiście, standardowo kolejnym krokiem jest wyznaczenie pochodnych względem \mathbf{v} oraz \mathbf{w} :

Taki zapis można również przedstawić w sformułowaniu lagranżianu:

$$\min_{\mathbf{a}, \mathbf{b}} \max_{\lambda_1, \lambda_2} \sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{y}_i \mathbf{w} - \frac{\lambda_1}{2} \left(\sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{x}_i \mathbf{v} - N \right) - \frac{\lambda_2}{2} \left(\sum_i \mathbf{w}^T \mathbf{y}_i^T \mathbf{y}_i \mathbf{w} - N \right)$$

Oczywiście, standardowo kolejnym krokiem jest wyznaczenie pochodnych względem \mathbf{v} oraz \mathbf{w} :

$$\begin{aligned} \sum_i \mathbf{x}_i^T \mathbf{y}_i \mathbf{w} - \lambda_1 \sum_i \mathbf{x}_i^T \mathbf{x}_i \mathbf{v} \\ \sum_i \mathbf{y}_i^T \mathbf{x}_i \mathbf{v} - \lambda_2 \sum_i \mathbf{y}_i^T \mathbf{y}_i \mathbf{w} \end{aligned}$$

Taki zapis można również przedstawić w sformułowaniu lagranżianu:

$$\min_{\mathbf{a}, \mathbf{b}} \max_{\lambda_1, \lambda_2} \sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{y}_i \mathbf{w} - \frac{\lambda_1}{2} \left(\sum_i \mathbf{v}^T \mathbf{x}_i^T \mathbf{x}_i \mathbf{v} - N \right) - \frac{\lambda_2}{2} \left(\sum_i \mathbf{w}^T \mathbf{y}_i^T \mathbf{y}_i \mathbf{w} - N \right)$$

Oczywiście, standardowo kolejnym krokiem jest wyznaczenie pochodnych względem \mathbf{v} oraz \mathbf{w} :

$$\begin{aligned} \sum_i \mathbf{x}_i^T \mathbf{y}_i \mathbf{w} - \lambda_1 \sum_i \mathbf{x}_i^T \mathbf{x}_i \mathbf{v} \\ \sum_i \mathbf{y}_i^T \mathbf{x}_i \mathbf{v} - \lambda_2 \sum_i \mathbf{y}_i^T \mathbf{y}_i \mathbf{w} \end{aligned}$$

Jeśli byśmy pomnożyli pierwsze równanie przez \mathbf{v}^T , a drugie przez \mathbf{w}^T i odjęli je stronami, to otrzymalibyśmy w rezultacie, że $\lambda_1 = \lambda_2 \equiv \lambda$.

Dla “uproszczenia życia” można wprowadzić następujący zapis:

Dla “uproszczenia życia” można wprowadzić następujący zapis:

$$\mathbf{S}_{xy} = \sum_i \mathbf{x}_i^T \mathbf{y}_i \quad \mathbf{S}_{xx} = \sum_i \mathbf{x}_i^T \mathbf{x}_i \quad \mathbf{S}_{yy} = \sum_i \mathbf{y}_i^T \mathbf{y}_i$$

Dla “uproszczenia życia” można wprowadzić następujący zapis:

$$\mathbf{S}_{xy} = \sum_i \mathbf{x}_i^T \mathbf{y}_i \quad \mathbf{S}_{xx} = \sum_i \mathbf{x}_i^T \mathbf{x}_i \quad \mathbf{S}_{yy} = \sum_i \mathbf{y}_i^T \mathbf{y}_i$$

co pozwala zapisać równania jako

Dla “uproszczenia życia” można wprowadzić następujący zapis:

$$\mathbf{S}_{xy} = \sum_i \mathbf{x}_i^T \mathbf{y}_i \quad \mathbf{S}_{xx} = \sum_i \mathbf{x}_i^T \mathbf{x}_i \quad \mathbf{S}_{yy} = \sum_i \mathbf{y}_i^T \mathbf{y}_i$$

co pozwala zapisać równania jako

$$\begin{aligned} \mathbf{S}_{xy} \mathbf{w} &= \lambda \mathbf{S}_{xx} \mathbf{v} \\ \mathbf{S}_{xy} \mathbf{v} &= \lambda \mathbf{S}_{yy} \mathbf{w} \end{aligned}$$

Dla “uproszczenia życia” można wprowadzić następujący zapis:

$$\mathbf{S}_{xy} = \sum_i \mathbf{x}_i^T \mathbf{y}_i \quad \mathbf{S}_{xx} = \sum_i \mathbf{x}_i^T \mathbf{x}_i \quad \mathbf{S}_{yy} = \sum_i \mathbf{y}_i^T \mathbf{y}_i$$

co pozwala zapisać równania jako

$$\begin{aligned} \mathbf{S}_{xy} \mathbf{w} &= \lambda \mathbf{S}_{xx} \mathbf{v} \\ \mathbf{S}_{xy} \mathbf{v} &= \lambda \mathbf{S}_{yy} \mathbf{w} \end{aligned}$$

Przekształcając pierwsze równanie do postaci $\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{w} = \lambda \mathbf{v}$, a następnie mnożąc je stronami przez \mathbf{S}_{xy} i porównując z drugim równaniem (i robiąc podobne przejście dla drugiego równania) otrzymujemy

Dla “uproszczenia życia” można wprowadzić następujący zapis:

$$\mathbf{S}_{xy} = \sum_i \mathbf{x}_i^T \mathbf{y}_i \quad \mathbf{S}_{xx} = \sum_i \mathbf{x}_i^T \mathbf{x}_i \quad \mathbf{S}_{yy} = \sum_i \mathbf{y}_i^T \mathbf{y}_i$$

co pozwala zapisać równania jako

$$\begin{aligned} \mathbf{S}_{xy} \mathbf{w} &= \lambda \mathbf{S}_{xx} \mathbf{v} \\ \mathbf{S}_{xy} \mathbf{v} &= \lambda \mathbf{S}_{yy} \mathbf{w} \end{aligned}$$

Przekształcając pierwsze równanie do postaci $\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{w} = \lambda \mathbf{v}$, a następnie mnożąc je stronami przez \mathbf{S}_{xy} i porównując z drugim równaniem (i robiąc podobne przejście dla drugiego równania) otrzymujemy

$$\begin{aligned} \mathbf{S}_{yy}^{-1} \mathbf{S}_{xy} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{w} &= \lambda^2 \mathbf{w} \\ \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{xy} \mathbf{v} &= \lambda^2 \mathbf{v} \end{aligned}$$

Dla “uproszczenia życia” można wprowadzić następujący zapis:

$$\mathbf{S}_{xy} = \sum_i \mathbf{x}_i^T \mathbf{y}_i \quad \mathbf{S}_{xx} = \sum_i \mathbf{x}_i^T \mathbf{x}_i \quad \mathbf{S}_{yy} = \sum_i \mathbf{y}_i^T \mathbf{y}_i$$

co pozwala zapisać równania jako

$$\begin{aligned} \mathbf{S}_{xy} \mathbf{w} &= \lambda \mathbf{S}_{xx} \mathbf{v} \\ \mathbf{S}_{xy} \mathbf{v} &= \lambda \mathbf{S}_{yy} \mathbf{w} \end{aligned}$$

Przekształcając pierwsze równanie do postaci $\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{w} = \lambda \mathbf{v}$, a następnie mnożąc je stronami przez \mathbf{S}_{xy} i porównując z drugim równaniem (i robiąc podobne przejście dla drugiego równania) otrzymujemy

$$\begin{aligned} \mathbf{S}_{yy}^{-1} \mathbf{S}_{xy} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{w} &= \lambda^2 \mathbf{w} \\ \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{xy} \mathbf{v} &= \lambda^2 \mathbf{v} \end{aligned}$$

Finalnie:

- wektory kanoniczne (\mathbf{v} , \mathbf{w}) są wektorami własnymi odpowiadającymi r największym wartościom własnym $1 \geq \lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_r^2 \geq 0$,

Finalnie:

- wektory kanoniczne (\mathbf{v}, \mathbf{w}) są wektorami własnymi odpowiadającymi r największym wartościom własnym $1 \geq \lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_r^2 \geq 0$,
- wartości własne są równe współczynnikom korelacji podniesionym do kwadratu,

Finalnie:

- wektory kanoniczne (\mathbf{v} , \mathbf{w}) są wektorami własnymi odpowiadającymi r największym wartościom własnym $1 \geq \lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_r^2 \geq 0$,
- wartości własne są równe współczynnikom korelacji podniesionym do kwadratu,
- kolejne pary (\mathbf{v}_j , \mathbf{w}_j) są do siebie ortogonalne.

Podobnie jak w przypadku SVM możemy się pokusić o wprowadzenie zależności nieliniowych, opisujących dane, tzn. wykorzystać koncepcje jąder. W tym celu przedstawiamy wektory kanoniczne (\mathbf{v} , \mathbf{w}) jako kombinacje liniowe

Podobnie jak w przypadku SVM możemy się pokusić o wprowadzenie zależności nieliniowych, opisujących dane, tzn. wykorzystac koncepcje jąder. W tym celu przedstawiamy wektory kanoniczne (\mathbf{v} , \mathbf{w}) jako kombinacje liniowe

$$\mathbf{v} = \sum_i \Phi(\mathbf{x}_i) \alpha_i \quad \mathbf{w} = \sum_i \Phi(\mathbf{y}_i) \beta_i ,$$

Podobnie jak w przypadku SVM możemy się pokusić o wprowadzenie zależności nieliniowych, opisujących dane, tzn. wykorzystac koncepcje jąder. W tym celu przedstawiamy wektory kanoniczne (\mathbf{v} , \mathbf{w}) jako kombinacje liniowe

$$\mathbf{v} = \sum_i \Phi(\mathbf{x}_i) \alpha_i \quad \mathbf{w} = \sum_i \Phi(\mathbf{y}_i) \beta_i ,$$

gdzie α_i i β_i to współczynniki rozwinięcia.

Podobnie jak w przypadku SVM możemy się pokusić o wprowadzenie zależności nieliniowych, opisujących dane, tzn. wykorzystac koncepcje jąder. W tym celu przedstawiamy wektory kanoniczne (\mathbf{v} , \mathbf{w}) jako kombinacje liniowe

$$\mathbf{v} = \sum_i \Phi(\mathbf{x}_i) \alpha_i \quad \mathbf{w} = \sum_i \Phi(\mathbf{y}_i) \beta_i ,$$

gdzie α_i i β_i to współczynniki rozwinięcia.

Podobnie jak w przypadku liniowym możemy taki układ zapisać jako funkcję Lagrange'a:

$$L = \alpha^T \mathbf{K}_x \mathbf{K}_y \beta - \frac{\lambda_1}{2} (\alpha^T \mathbf{K}_x^2 \alpha - N) - \frac{\lambda_2}{2} (\beta^T \mathbf{K}_y^2 \beta - N) ,$$

Podobnie jak w przypadku SVM możemy się pokusić o wprowadzenie zależności nieliniowych, opisujących dane, tzn. wykorzystac koncepcje jąder. W tym celu przedstawiamy wektory kanoniczne (\mathbf{v}, \mathbf{w}) jako kombinacje liniowe

$$\mathbf{v} = \sum_i \Phi(\mathbf{x}_i) \alpha_i \quad \mathbf{w} = \sum_i \Phi(\mathbf{y}_i) \beta_i ,$$

gdzie α_i i β_i to współczynniki rozwinięcia.

Podobnie jak w przypadku liniowym możemy taki układ zapisać jako funkcję Lagrange'a:

$$L = \alpha^T \mathbf{K}_x \mathbf{K}_y \beta - \frac{\lambda_1}{2} (\alpha^T \mathbf{K}_x^2 \alpha - N) - \frac{\lambda_1}{2} (\beta^T \mathbf{K}_y^2 \beta - N) ,$$

,przy czym współczynniki $\alpha, \beta \in R^N$, natomiast

$$\mathbf{K}_x = \sum_i \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_i),$$

Podobnie jak w przypadku SVM możemy się pokusić o wprowadzenie zależności nieliniowych, opisujących dane, tzn. wykorzystac koncepcje jąder. W tym celu przedstawiamy wektory kanoniczne (\mathbf{v}, \mathbf{w}) jako kombinacje liniowe

$$\mathbf{v} = \sum_i \Phi(\mathbf{x}_i) \alpha_i \quad \mathbf{w} = \sum_i \Phi(\mathbf{y}_i) \beta_i ,$$

gdzie α_i i β_i to współczynniki rozwinięcia.

Podobnie jak w przypadku liniowym możemy taki układ zapisać jako funkcję Lagrange'a:

$$L = \alpha^T \mathbf{K}_x \mathbf{K}_y \beta - \frac{\lambda_1}{2} (\alpha^T \mathbf{K}_x^2 \alpha - N) - \frac{\lambda_2}{2} (\beta^T \mathbf{K}_y^2 \beta - N) ,$$

,przy czym współczynniki $\alpha, \beta \in R^N$, natomiast

$$\mathbf{K}_x = \sum_i \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_i),$$

jest macierzą Grama i tworzy się ją jako $(\mathbf{K}_x)_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$.

Rozważania podobne do tych w przypadku liniowej CCA prowadzą do układu równań

Rozważania podobne do tych w przypadku liniowej CCA prowadzą do układu równań

$$\begin{aligned}K_x K_y \beta &= \lambda K_x^2 \alpha \\K_y K_x \alpha &= \lambda K_y^2 \beta \ .\end{aligned}$$

Rozważania podobne do tych w przypadku liniowej CCA prowadzą do układu równań

$$\begin{aligned}K_x K_y \beta &= \lambda K_x^2 \alpha \\ K_y K_x \alpha &= \lambda K_y^2 \beta\end{aligned}$$

z którego możemy wyznaczyć współczynniki α oraz β .

Rozważania podobne do tych w przypadku liniowej CCA prowadzą do układu równań

$$\begin{aligned} K_X K_Y \beta &= \lambda K_X^2 \alpha \\ K_Y K_X \alpha &= \lambda K_Y^2 \beta \end{aligned}$$

z którego możemy wyznaczyć współczynniki α oraz β .

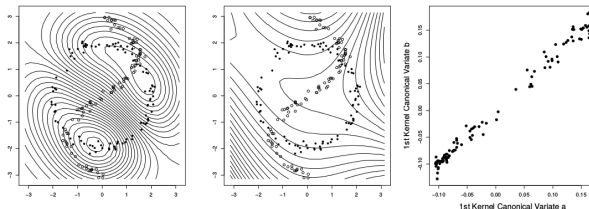


Figure 3: Kernel canonical correlation example. The data consists of two sets of 100 points each. For \mathbf{X} the points are lying on a circle (solid points) while \mathbf{Y} (circles) describe a sine curve (points correspond by arclength). For \mathbf{X} we used a RBF kernel ($\sigma = 1$) and for \mathbf{Y} a homogeneous polynomial kernel of degree ($d = 2$). The lines plotted describe regions of equal score on the first canonical vectors, which can be thought of as orthogonal (see Schölkopf et al. (1998)). This is shown for $\mathbf{v}_1 \in \mathcal{L}\{\Phi'_X\}$ (upper) and for $\mathbf{w}_1 \in \mathcal{L}\{\Phi'_Y\}$ (middle). The bottom plot shows the first pair of kernel canonical variates ($\mathbf{a}_1, \mathbf{b}_1$) showing that $\langle \phi(\mathbf{x}_i), \mathbf{v}_1 \rangle_{\mathcal{F}}$ and $\langle \phi(\mathbf{y}_i), \mathbf{w}_1 \rangle_{\mathcal{F}}$ are highly correlated for $i = 1, \dots, m$.

Przypomnienie: dla zbioru scentrowanych obserwacji \mathbf{x}_i , $i = 1, \dots, N$, metoda składowych głównych sprowadza się do diagonalizacji macierzy kowariancji

Przypomnienie: dla zbioru scentrowanych obserwacji \mathbf{x}_i , $i = 1, \dots, N$, metoda składowych głównych sprowadza się do diagonalizacji macierzy kowariancji

$$\mathbf{S} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \mathbf{x}_j^T,$$

Przypomnienie: dla zbioru scentrowanych obserwacji \mathbf{x}_i , $i = 1, \dots, N$, metoda składowych głównych sprowadza się do diagonalizacji macierzy kowariancji

$$\mathbf{S} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \mathbf{x}_j^T,$$

za pomocą rozwiązania zagdanienia własnego

$$\lambda \mathbf{v} = \mathbf{S} \mathbf{v},$$

Przypomnienie: dla zbioru scentrowanych obserwacji \mathbf{x}_i , $i = 1, \dots, N$, metoda składowych głównych sprowadza się do diagonalizacji macierzy kowariancji

$$\mathbf{S} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \mathbf{x}_j^T,$$

za pomocą rozwiązania zagdanienia własnego

$$\lambda \mathbf{v} = \mathbf{S} \mathbf{v},$$

Ponieważ wszystkie rozwiązania \mathbf{v} z $\lambda \neq 0$ są z przestrzeni rozpiętej na $\mathbf{x}_1, \dots, \mathbf{x}_N$ powyższe równanie można zapisać jako

Przypomnienie: dla zbioru scentrowanych obserwacji \mathbf{x}_i , $i = 1, \dots, N$, metoda składowych głównych sprowadza się do diagonalizacji macierzy kowariancji

$$\mathbf{S} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \mathbf{x}_j^T,$$

za pomocą rozwiązania zagadnienia własnego

$$\lambda \mathbf{v} = \mathbf{S} \mathbf{v},$$

Ponieważ wszystkie rozwiązania \mathbf{v} z $\lambda \neq 0$ są z przestrzeni rozpiętej na $\mathbf{x}_1, \dots, \mathbf{x}_N$ powyższe równanie można zapisać jako

$$\lambda \mathbf{x}_k \cdot \mathbf{v} = \mathbf{x}_k \cdot \mathbf{S} \mathbf{v} \quad \text{dla} \quad k = 1, \dots, N$$

Przypomnienie: dla zbioru scentrowanych obserwacji \mathbf{x}_i , $i = 1, \dots, N$, metoda składowych głównych sprowadza się do diagonalizacji macierzy kowariancji

$$\mathbf{S} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \mathbf{x}_j^T,$$

za pomocą rozwiązania zagadnienia własnego

$$\lambda \mathbf{v} = \mathbf{S} \mathbf{v},$$

Ponieważ wszystkie rozwiązania \mathbf{v} z $\lambda \neq 0$ są z przestrzeni rozpiętej na $\mathbf{x}_1, \dots, \mathbf{x}_N$ powyższe równanie można zapisać jako

$$\lambda \mathbf{x}_k \cdot \mathbf{v} = \mathbf{x}_k \cdot \mathbf{S} \mathbf{v} \quad \text{dla} \quad k = 1, \dots, N$$

Przejdźmy teraz do “wzbogaconej” przestrzeni $\Psi(\mathbf{x})$, w której analogiczna macierz kowariancji $\hat{\mathbf{S}}$ będzie miała postać

Przejdźmy teraz do “wzbogaconej” przestrzeni $\Psi(\mathbf{x})$, w której analogiczna macierz kowariancji $\hat{\mathbf{S}}$ będzie miała postać

$$\hat{\mathbf{S}} = \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T,$$

Przejdźmy teraz do “wzbogaconej” przestrzeni $\Psi(\mathbf{x})$, w której analogiczna macierz kowariancji $\hat{\mathbf{S}}$ będzie miała postać

$$\hat{\mathbf{S}} = \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T,$$

Podobnie jak poprzednio, szukamy wartości własnych λ oraz wektorów własnych \mathbf{V} dla

$$\lambda \mathbf{V} = \hat{\mathbf{S}} \mathbf{V},$$

Przejdźmy teraz do “wzbogaconej” przestrzeni $\Psi(\mathbf{x})$, w której analogiczna macierz kowariancji $\hat{\mathbf{S}}$ będzie miała postać

$$\hat{\mathbf{S}} = \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T,$$

Podobnie jak poprzednio, szukamy wartości własnych λ oraz wektorów własnych \mathbf{V} dla

$$\lambda \mathbf{V} = \hat{\mathbf{S}} \mathbf{V},$$

co możemy również zapisać jako

$$\lambda \Phi(\mathbf{x}_k) \cdot \mathbf{V} = \Phi(\mathbf{x}_k) \cdot \hat{\mathbf{S}} \mathbf{V} \quad \text{dla} \quad k = 1, \dots, N$$

Przejdźmy teraz do “wzbogaconej” przestrzeni $\Psi(\mathbf{x})$, w której analogiczna macierz kowariancji $\hat{\mathbf{S}}$ będzie miała postać

$$\hat{\mathbf{S}} = \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T,$$

Podobnie jak poprzednio, szukamy wartości własnych λ oraz wektorów własnych \mathbf{V} dla

$$\lambda \mathbf{V} = \hat{\mathbf{S}} \mathbf{V},$$

co możemy również zapisać jako

$$\lambda \Phi(\mathbf{x}_k) \cdot \mathbf{V} = \Phi(\mathbf{x}_k) \cdot \hat{\mathbf{S}} \mathbf{V} \quad \text{dla} \quad k = 1, \dots, N$$

Ponieważ wektor \mathbf{V} rozpina się w przestrzeni $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_N)$, można go zapisać w postaci

$$\mathbf{V} = \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_i),$$

W efekcie rozważamy równanie

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$.

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j),$$

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j),$$

co prowadzi do równania

$$N\lambda\alpha = \mathbf{K}\alpha.$$

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j),$$

co prowadzi do równania

$$N\lambda\alpha = \mathbf{K}\alpha.$$

Dodatkowo normalizujemy α^k , żądając aby $\mathbf{V}^k \cdot \mathbf{V}^k = 1$.

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j),$$

co prowadzi do równania

$$N\lambda\alpha = \mathbf{K}\alpha.$$

Dodatkowo normalizujemy α^k , żądając aby $\mathbf{V}^k \cdot \mathbf{V}^k = 1$. Aby wyznaczyć składowe główne, musimy policzyć rzuty na wektory własne \mathbf{V}^k , tzn:

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j),$$

co prowadzi do równania

$$N\lambda\alpha = \mathbf{K}\alpha.$$

Dodatkowo normalizujemy α^k , żądając aby $\mathbf{V}^k \cdot \mathbf{V}^k = 1$. Aby wyznaczyć składowe główne, musimy policzyć rzuty na wektory własne \mathbf{V}^k , tzn:

$$\mathbf{V}^k \cdot \Phi(\mathbf{x}) = \sum_i^N \alpha_i^k \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}).$$

W efekcie rozważamy równanie

$$\lambda \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i) = \frac{1}{M} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_k) \cdot \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i),$$

dla $k = 1, \dots, N$. Podobnie jak w przypadku kCCA możemy zdefiniować macierz Grama \mathbf{K}

$$\mathbf{K}_{ij} = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j),$$

co prowadzi do równania

$$N\lambda\alpha = \mathbf{K}\alpha.$$

Dodatkowo normalizujemy α^k , żądając aby $\mathbf{V}^k \cdot \mathbf{V}^k = 1$. Aby wyznaczyć składowe główne, musimy policzyć rzuty na wektory własne \mathbf{V}^k , tzn:

$$\mathbf{V}^k \cdot \Phi(\mathbf{x}) = \sum_i^N \alpha_i^k \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}).$$