Estrategias de representación utilizando acople entre espacios → Metric Learning.

1. Alineamiento de kernel centralizado.

$X \in \mathcal{X}$; $Y \in \mathcal{Y}$

$\varphi_x : \mathcal{X} \to \mathcal{H}_x$; $\varphi_y : \mathcal{Y} \to \mathcal{H}_y$.

Ej: aprendizaje de métrica → relación con modelo lineal $Z = XA$; $\mathcal{X} \subseteq \mathbb{R}^P$; $A \in \mathbb{R}^{P \times Q}$

$P \leq Q$; $Z \in \mathbb{R}^{N \times Q}$; $X \in \mathbb{R}^{N \times P}$

$$d(z_n, z_m) = d(x_n A, x_m A) = \| x_n A - x_m A \|^2 = \langle x_n A - x_m A, x_n A - x_m A \rangle$$

$$d(z_n, z_m) = (x_n - x_m) A A^T (x_n - x_m)^T$$

$$[K_z(A)]_{n,m} = K_G(x_n - x_m \mid \Sigma^{-1} = AA^T).$$

$$[K_y]_{n,m} = \kappa_y(y_n, y_m)$$

$$A^* = \arg\max_A \hat{\rho}(K_z(A), K_y) = \frac{\langle \bar{K}_z(A), \bar{K}_y \rangle_F}{\| \bar{K}_z(A) \|_F \| \bar{K}_y \|_F}$$

$$\bar{K} = HKH \; ; \; H = I - \frac{1}{N} 1_N 1_N^T.$$

CKA Metric Learning → Deep Learning.

input $= X$

↓

Dense Layer $= XW$
bias = False
Activation = Linear
Units $= Q$ → agregar regularizador/restricción.

↓ y_{pred}

Output
Custom loss $= \rho(K(y_{pred}), K(y_{true}))$

$K(y_{pred}) = k_G(y_{pred_n}, y_{pred_m} \mid \Sigma = I)$

$K(y_{true}) = k_G(y_{true_n}, y_{true_m} \mid \Sigma = \sigma_y I)$

$\sigma_y \in \mathbb{R}^+ \to$ según tarea.

Ej: clasificación $\sigma_y \to 0$.

2. Canonical Correlation → Correlación Canónica.

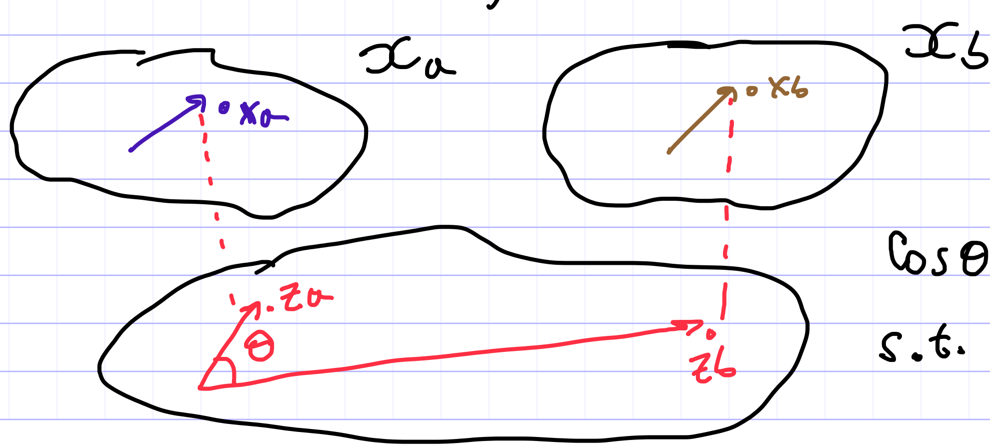
→ Dos vistas de los datos → Multi-view

CCA: Canonical Correlation Analysis

↳ Acople de espacios. $\quad X_a \in \mathcal{X}_a$; $X_b \in \mathcal{X}_b$

$\mathcal{X}_a \subseteq \mathbb{R}^P$; $\mathcal{X}_b \subseteq \mathbb{R}^Q$

$\mathcal{D} = \{ X_a \in \mathbb{R}^{N \times P}, X_b \in \mathbb{R}^{N \times Q} \}$



$\cos\theta = \langle z_a, z_b \rangle$

s.t. $\|z_a\|_2 = \|z_b\|_2 = 1$.

Formulación ortogonal CCA.

$\mathcal{D} = \{ \mathbb{X}_a \in \mathbb{R}^{N\times P}, \mathbb{X}_b \in \mathbb{R}^{N\times Q} \}$

$\mathbb{Z}_a = \mathbb{X}_a \mathbb{W}_a \in \mathbb{R}^{N\times L}$; $\mathbb{Z}_b = \mathbb{X}_b \mathbb{W}_b \in \mathbb{R}^{N\times L}$

$L = \min(P, Q)$.

$\mathbb{W}_a = [\mathbb{w}_a^1, \dots, \mathbb{w}_a^L] \in \mathbb{R}^{P\times L}$; $\mathbb{W}_b = [\mathbb{w}_b^1, \dots, \mathbb{w}_b^L] \in \mathbb{R}^{Q\times L}$

Para r-th par $\mathbb{w}_a^r \in \mathbb{R}^P, \mathbb{w}_b^r \in \mathbb{R}^Q$; $r \in \{1, \dots, L\}$

$$\cos(\theta_r) = \max_{\mathbb{w}_a^r, \mathbb{w}_b^r} \langle \mathbb{z}_a^r, \mathbb{z}_b^r \rangle$$

$$\text{s.t.} \quad \langle \mathbb{z}_a^r, \mathbb{z}_a^j \rangle = \begin{cases} 1 & r = j \\ 0 & r \neq j \end{cases}$$

$$\langle \mathbb{z}_b^r, \mathbb{z}_b^j \rangle = \begin{cases} 1 & r = j \\ 0 & r \neq j \end{cases}$$

Solución analítica:

Dado un par de vectores $\mathbb{w}_a \in \mathbb{R}^P$; $\mathbb{w}_b \in \mathbb{R}^Q$.

$\mathbb{z}_a = \mathbb{X}_a \mathbb{w}_a \in \mathbb{R}^N$; $\mathbb{z}_b = \mathbb{X}_b \mathbb{w}_b \in \mathbb{R}^N$

$$\cos(\theta) = \max_{\mathbb{w}_a, \mathbb{w}_b} \langle \mathbb{z}_a, \mathbb{z}_b \rangle = \max_{\mathbb{w}_a, \mathbb{w}_b} \langle \mathbb{X}_a \mathbb{w}_a, \mathbb{X}_b \mathbb{w}_b \rangle$$

$$\cos(\theta) = \max_{w_a, w_b} w_a^T X_a^T X_b w_b$$

$$\text{s.t. } \|z_a\|_2^2 = \langle X_a w_a, X_a w_a \rangle = 1$$

$$\|z_a\|_2^2 = w_a^T X_a^T X_a w_a = 1.$$

$$\|z_b\|_2^2 = w_b^T X_b^T X_b w_b = 1$$

Lagrangiano:

$$L = w_a^T C_{ab} w_b - \rho_1 (w_a^T C_{aa} w_a - 1) - \rho_2 (w_b^T C_{bb} w_b - 1)$$

$C_{ab} = X_a^T X_b \in \mathbb{R}^{P \times Q}$; $C_{aa} = X_a^T X_a \in \mathbb{R}^{P \times P}$

$C_{bb} = X_b^T X_b \in \mathbb{R}^{Q \times Q}$.

$$\frac{\partial L}{\partial w_a} = C_{ab} w_b - 2\rho_1 C_{aa} w_a = 0$$

$$\frac{\partial L}{\partial w_b} = C_{ab}^T w_a - 2\rho_2 C_{bb} w_b = 0; \quad C_{ab}^T = X_b^T X_a = C_{ba}$$

$$w_a^T C_{ab} w_b - 2\rho_1 w_a^T C_{aa} w_a = w_a^T C_{ab} w_b - 2\rho_1 \|z_a\|_2^2 = 0$$

$$w_b^T C_{ab}^T w_a - 2\rho_2 w_b^T C_{bb} w_b = w_b^T C_{ab}^T w_a - 2\rho_2 \|z_b\|_2^2 = 0$$

$$w_a^T C_{ab} w_b = w_b^T C_{ab}^T w_a = 2\rho_1 = 2\rho_2$$

$$\rho_1 = \rho_2 = \rho$$

$$\frac{\partial L}{\partial W_a} = C_{ab} W_b - 2\rho_1 C_{aa} W_a = 0$$

$$2\rho_1 = \tilde{\rho} \rightarrow \qquad \tilde{\rho}\, C_{aa} W_a = C_{ab} W_b$$

$$\boxed{W_a = \frac{C_{aa}^{-1}}{\tilde{\rho}} C_{ab} W_b}$$

$$\frac{\partial L}{\partial W_b} = C_{ab}^T W_a - \tilde{\rho}\, C_{bb} W_b = 0$$

$$\frac{C_{ab}^T C_{aa}^{-1} C_{ab} W_b}{\tilde{\rho}} - \tilde{\rho}\, C_{bb} W_b = 0$$

$$C_{ab}^T C_{aa}^{-1} C_{ab} W_b = \tilde{\rho}^2 C_{bb} W_b$$

$$C_{bb}^{-1} C_{ab}^T C_{aa}^{-1} C_{ab} W_b = \tilde{\rho}^2 W_b.$$

$$\boxed{A W_b = \lambda W_b \longrightarrow \text{eig problem}}$$

Resolver por eig. asegura ortogonalidad:

$W_b = \text{eig}(A) \rightarrow$ encontrar L eigenvalues
eigenvectors

$$W_a = \frac{C_{aa}^{-1}}{\sqrt{\lambda}} C_{ab} W_b.$$

$$\langle W_b^r, W_b^j \rangle = \begin{cases} 1 & r = j \\ 0 & r \neq j \end{cases} ; \quad \langle W_a^r, W_a^j \rangle = \begin{cases} 1 & r = j \\ 0 & r \neq j \end{cases}$$

CCA regularizado.

$$\cos(\theta) = \max_{w_a \in \mathbb{R}^p,\, w_b \in \mathbb{R}^q} w_a^T C_{ab} w_b$$

$$\|w_a\|^2_{C_{aa}+\lambda_a I} = w_a^T (C_{aa}+\lambda_a I) w_a$$

$$\|w_b\|^2_{C_{bb}+\lambda_b I} = w_b^T (C_{bb}+\lambda_b I) w_b$$

Por ende el problema de valores propios se reescribe como:

$$(C_{bb}+\lambda_b I)^{-1} C_{ab}^T (C_{aa}+\lambda_a I)^{-1} C_{ab} w_b = \tilde{\rho}^2 w_b.$$

$$w_a = \underbrace{(C_{aa}+\lambda_a I)}_{\tilde{\rho}}{}^{-1} C_{ab} w_b.$$

Recuerda que regularizar la matriz se entiende como un filtro espectral desde las bases de reconstrucción del espacio Hilbert.

Ej: $X^T X \in \mathbb{R}^{p \times p} = V \Delta V^T$; $\quad VV^T = I$

$\quad V^T V = I.$

$X X^T \in \mathbb{R}^{N \times N} = U \tilde{\Delta} U^T$

$\quad \Delta = \operatorname{diag}(\rho_1, \dots, \rho_p)$

$X = USV^T \rightarrow \quad SS^* = \Delta \qquad UU^T = I, \; U^T U = I$

$s_j^2 = \rho_j \qquad \tilde{\Delta} = \operatorname{diag}(\rho_1, \dots, \rho_N)$

valor singular $\qquad$ valor propio

$$(X^T X)^{-1} = (V S^* S\ V^T)^{-1} \longrightarrow V (S^* S)^{-1} V^T$$

TAREA: Estudiar teorema representación espectral

$$(X^T X)^{-1} = V (S^2)^{-1} V^T = V S^{-2} V^T = V \begin{bmatrix} \frac{1}{s_1^2} & & \\ & \ddots & \\ & & \frac{1}{s_p^2} \end{bmatrix} V^T$$

Regularizando:

$$(X^T X + \lambda I)^{-1} = (V S^* S V^T + \lambda I) = (V (S^* S + \lambda I) V^T)^{-1}$$

$$= V (S^2 + \lambda I)^{-1} V^T = V \begin{bmatrix} \frac{1}{s_1^2+\lambda} & & \\ & \ddots & \\ & & \frac{1}{s_p^2+\lambda} \end{bmatrix} V^T .$$

NOTA: Recuerde que la inversa en deep learning la evitamos por gradiente descendiente; pero el concepto de regularización (filtrado) se mantiene sobre los pesos de las neuronas.

Kernel CCA: $x_a \in \mathcal{X}_a$; $x_b \in \mathcal{X}_b$

$$\kappa_a : \mathcal{X}_a \times \mathcal{X}_a \to \mathbb{R}; \; \kappa_b : \mathcal{X}_b \times \mathcal{X}_b \to \mathbb{R}.$$

$$z_a = K_a \alpha_a ; \quad z_b = K_b \alpha_b ; \quad K_a \in \mathbb{R}^{N \times N} ; \quad K_b \in \mathbb{R}^{N \times N}$$

$$\varphi_a : \mathcal{X}_a \to \mathcal{H}_a ; \quad \varphi_b : \mathcal{X}_b \to \mathcal{H}_b$$

$$\cos(\theta) = \max_{\alpha_a, \alpha_b \in \mathbb{R}^N} \langle z_a, z_b \rangle = \langle K_a \alpha_a, K_b \alpha_b \rangle$$

$$= \alpha_a^T K_a^T K_b \alpha_b$$

$$\text{s.t. } \|z_a\|_2 = \|z_b\|_2 = 1$$

$$\sqrt{\alpha_a^T K_a^T K_a \alpha_a} = \sqrt{\alpha_b^T K_b^T K_b \alpha_b} = 1$$

$$L = \alpha_a^T K_a^T K_b \alpha_b - \rho_1 (\alpha_a^T K_a^2 \alpha_a - 1) - \rho_2 (\alpha_b^T K_b^2 \alpha_b - 1)$$

$$K_a^T K_a = K_a K_a^T = K_a^2$$

$$\frac{\partial L}{\partial \alpha_a} = K_a^T K_b \alpha_b - 2 \rho_1 K_a^2 \alpha_a = 0$$

$$\frac{\partial L}{\partial \alpha_b} = K_b^T K_a \alpha_a - 2 \rho_2 K_b^2 \alpha_b = 0$$

$$\alpha_a^T K_a^T K_b \alpha_b - 2 \rho_1 \overbrace{\alpha_a^T K_a^2 \alpha_a}^{1} = 0$$

$$\alpha_b^T K_b^T K_a \alpha_a - 2 \rho_2 \underbrace{\alpha_b^T K_b^2 \alpha_b}_{1} = 0 \quad \longrightarrow \quad \rho_1 = \rho_2 .$$

$$K_a^T K_b \alpha_b - \tilde{\rho} K_a^2 \alpha_a = 0 \implies \alpha_a = \frac{K_a^{-1} K_a^{-1}}{\tilde{\rho}} K_a^T K_b \alpha_b$$

$$\boxed{\alpha_a = \frac{K_a^{-1}}{\tilde{\rho}} K_b \alpha_b}$$

$$\frac{K_b^T K_a K_a^{-1} K_b \alpha_b}{\tilde{\rho}} - \tilde{\rho} K_b^2 \alpha_b = 0$$

$$K_b^2 \alpha_b - \tilde{\rho}^2 K_b^2 \alpha_b = 0 \rightarrow K_b^2 \alpha_b = \tilde{\rho}^2 K_b^2 \alpha_b$$

$$I \alpha_b = \rho^2 \alpha_b \rightarrow \text{Si } K_b \text{ es invertible}$$

$$KCCA \rightarrow I.$$

Se requiere de regularización.!

$$\cos(\theta) = \max_{\alpha_a, \alpha_b \in \mathbb{R}^N} \langle z_a, z_b \rangle = \langle K_a \alpha_a, K_b \alpha_b \rangle$$

$$\|z_a\|^2_{K_a + \lambda_a I} = \langle (K_a + \lambda_a I)\alpha_a, (K_a + \lambda_a I)\alpha_a \rangle = \alpha_a^T (K_a + \lambda_a I)^2 \alpha_a = 1$$

$$\|z_b\|^2_{K_b + \lambda_b I} = \alpha_b^T (K_b + \lambda_b I)^2 \alpha_b = 1$$

$$L = \alpha_a^T K_a^T K_b \alpha_b - \rho_1 (\alpha_a^T (K_a + \lambda_a I)^2 \alpha_a - 1) - \rho_2 (\alpha_b^T (K_b + \lambda_b I)^2 \alpha_b - 1)$$

$$\frac{\partial L}{\partial \alpha_a} = K_a^T K_b \alpha_b - 2\rho_1 (K_a + \lambda_a I)^2 \alpha_a = 0$$

$$\frac{\partial L}{\partial \alpha_b} = K_b^T K_a \alpha_a - 2\rho_2 (K_b + \lambda_b I)^2 \alpha_b = 0$$

$$\rho_1 = \rho_2 \ ;$$

$$\boxed{\alpha_a = \frac{1}{\tilde{\rho}} (K_a + \lambda_a I)^{-2} K_a^T K_b \alpha_b}$$

$$\frac{1}{\tilde{\rho}} K_b^T K_a \left(K_a + \lambda_a I\right)^{-2} K_a^T K_b \alpha_b - \tilde{\rho} \left(K_b + \lambda_b I\right)^{2} \alpha_b = 0$$

$$K_b K_a \left(K_a + \lambda_a I\right)^{-2} K_a K_b \alpha_b = \tilde{\rho}^2 \left(K_b + \lambda_b I\right)^{2} \alpha_b$$

$$\underbrace{\left(K_b + \lambda_b I\right)^{-2} K_b K_a \left(K_a + \lambda_a I\right)^{-2} K_a K_b}_{A} \alpha_b = \tilde{\rho}^2 \alpha_b$$

$$\boxed{A \alpha_b = \tilde{\rho}^2 \alpha_b}$$

TAREA: 1). Implementar KCCA en tf 2.0.

2). Implementar KCCA aprendiendo K_a y K_b como $K_G(\cdot \mid \Sigma^{-1} = WW^T)$