# CCA: all the math here

Guohao Dou

October 1, 2018

## 1 Formulation

CCA (Canonical Correlation Analysis) has the following training objective:

$$(w_1^*, w_2^*) = \text{argmax}_{w_1,w_2} \frac{w_1' \Sigma_{12} w_2}{\sqrt{w_1' \Sigma_{11} w_1 \cdot w_2' \Sigma_{22} w_2}}$$

Which could be easily turned into the following since we could scale $w_1, w_2$ however we want.

$$\begin{aligned}(w_1^*, w_2^*) &= \text{argmax}_{w_1,w_2} w_1' \Sigma_{12} w_2 \\ s.t.\ \ w_1' \Sigma_{11} w_1 &= 1; \ \ w_2' \Sigma_{22} w_2 = 1\end{aligned}$$

Also, we need some sort of orthogonality

$$w_1^{i'} \Sigma_{11} w_1^j = 0 \quad \forall i \neq j$$

$$w_2^{i'} \Sigma_{22} w_2^j = 0 \quad \forall i \neq j$$

where $w_1^i$ is the $i^{th}$ vector retrieved.

The orthogonality constraints are annoying, but luckily we can find all $w_1, w_2$ we care to find in one go with the orthogonality constraints preserved.

Say we want $k$ pairs of $w_1, w_2$.

Let's bundle the $k$ vectors of $w_1$ into columns of $A_1 = [w_1^{(1)}, w_1^{(2)}, ...w_1^{(k)}]$ and do the same for $A_2$.

The training objective could then be formulated as:

$$\begin{aligned}A_1^*, A_2^* &= \text{argmax } \text{tr}(A_1' \Sigma_{12} A_2) \\ &= \text{argmax} \sum_{i=1}^{k} w_1^{i'} \Sigma_{12} w_2^i \\ s.t.\ \ A_1' \Sigma_{11} A_1 &= I \\ A_2' \Sigma_{22} A_2 &= I\end{aligned}$$

And some really smart people showed that:

$$\begin{aligned}
T &\triangleq \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} \\
T &= UDV^T \\
A_1^* &= \Sigma_{11}^{-1/2}U_k \\
A_2^* &= \Sigma_{22}^{-1/2}V_k
\end{aligned}$$

where $U_k, V_k$ are top $k$ left and right singular vectors of $T$.
And to train a neural network, we could simply maximize the empirical correlation:

$$\max c\bar{o}rr(H_1, H_2) = \mathrm{tr}(\sqrt{T'T})$$

which is basically the trace norm of $T$.

Well, we are technically done. But importantly, we need to show *why* $\mathrm{tr}(\sqrt{T'T})$ is the empirical correlation at all: it doesn't look like one, does it?

To establish connection between the two, we just need to show that:

- $(A_1^* = \Sigma_{11}^{-1/2}U_k; A_2^* = \Sigma_{22}^{-1/2}V_k) \rightarrow \mathrm{maximized}(\mathrm{tr}(\sqrt{T'T}))$

Let's just set $k = o$, for cleaner math. Also, remember $T \overset{svd}{=} UDV^T$.

$$\begin{aligned}
\mathrm{tr}(A_1'\Sigma_{12}A_2) &= \mathrm{tr}(U^T\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}V) \\
&= \mathrm{tr}(U^TTV) \\
&= \mathrm{tr}(D) \\
&= \mathrm{tr}(\sqrt{T'T})
\end{aligned}$$