# Doing Gradient with Ledoit-Wolf a.k.a. Calculus 101

Guohao Dou

University of Illinois

*gdou2@illinois.edu*

November 11, 2018

# Extending DeepCCA Gradient - the original



Let $f$ be the loss function; $\hat{\Sigma}$ the estimated covariance matrix; $H_1, H_2$ the two modalities.

$$f = \mathrm{tr}(\sqrt{T^T T}), \text{ where } T = \hat{\Sigma}_{11}^{-1/2} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1/2}$$

From the original DCCA paper, we need to figure out the following to compute the gradient $\frac{\partial f}{\partial H_1}$.

$$\nabla_{11} = \frac{\partial f}{\partial \hat{\Sigma}_{11}}, \; \nabla_{12} = \frac{\partial f}{\partial \hat{\Sigma}_{12}},$$

$$\frac{\partial \hat{\Sigma}_{11}^{ab}}{\partial H_1}, \; \frac{\partial \hat{\Sigma}_{12}^{ab}}{\partial H_1}$$

# Extending DeepCCA Gradient - Ledoit

The Ledoit-Wolf estimator works as following

$$\hat{\Sigma} = (1 - s)\Sigma + s \cdot \frac{\mathrm{tr}(\Sigma)}{D} I$$

where $\Sigma$ is the "sample covariance matrix"(SCM) and $s$ is the shrinkage factor estimated from $\Sigma$. $D$ is the dimension of feature space. Note that there's some re-definition of symbols going on.

**Q**

How could we do gradient when we use Ledoit-Wolf instead of SCM to form $\hat{\Sigma}_{11}$?

**A**

$\frac{\partial f}{\partial \hat{\Sigma}_{11}^{ab}}$ is already provided. $\frac{\partial \Sigma_{11}^{cd}}{\partial H_1}$ is also provided. We just have to figure out the behaviour of $\frac{\partial \hat{\Sigma}_{11}^{ab}}{\partial \Sigma_{11}^{cd}}$

## Work out $\frac{\partial \hat{\Sigma}_{11}^{ab}}{\partial \Sigma_{11}^{cd}}$



$$\frac{\partial \hat{\Sigma}_{11}^{ab}}{\partial \Sigma_{11}^{cd}} = \left\{ \begin{array}{ll} \left\{ \begin{array}{ll} 1-s & \text{for } a=c, b=d \\ 0 & \text{otherwise} \end{array} \right\}, & \text{for } a \neq b \\ \left\{ \begin{array}{ll} 0 & \text{for } c \neq d \\ 1-s+s/D, & \text{for } a=c=d \\ s/D, & \text{otherwise} \end{array} \right\}, & \text{for } a=b \end{array} \right\}$$

$$= \delta_{ac} \cdot \delta_{bd} \cdot (1-s) + \delta_{ab} \cdot \delta_{cd} \cdot \frac{s}{D}$$

Doesn't look too bad, does it?

# Compute Gradient

$$
\begin{aligned}
\frac{\partial f}{\partial \Sigma_{11}^{cd}} &= \sum_{ab} \frac{\partial f}{\partial \hat{\Sigma}_{11}^{ab}} \cdot \frac{\partial \hat{\Sigma}_{11}^{ab}}{\partial \Sigma_{11}^{cd}} \\
&= (1-s) \cdot \frac{\partial f}{\partial \hat{\Sigma}_{11}^{cd}} + \left\{ \begin{array}{ll} 0 & \text{for } c \neq d \\ \frac{s}{D} \sum_a \frac{\partial f}{\partial \hat{\Sigma}_{11}^{aa}} & \text{for } c = d \end{array} \right\} \\
&= (1-s) \cdot \frac{\partial f}{\partial \hat{\Sigma}_{11}^{cd}} + \delta_{cd} \cdot \frac{s}{D} \cdot \mathrm{tr}(\nabla_{11}) \\
&= (1-s) \cdot \nabla_{11}^{cd} + \delta_{cd} \cdot \frac{s}{D} \cdot \mathrm{tr}(\nabla_{11})
\end{aligned}
$$

Then all we need to do is to form the old $\nabla_{11}$ and use it to form

$$
\nabla_{11}' = (1-s) \cdot \nabla_{11} + \frac{s}{D} \cdot \mathrm{tr}(\nabla_{11}) \cdot I
$$

# PyTorch Implementation

**End**