Distance Metric Learning Formula

Ye Jiatao

03/05/18

Because there is no detailed derivation processes for the first derivative and second derivative used in Newton-Raphson update rule. Here we give a little more information about the how to get first derivative and Hessian Matrix.

$$g(A) = g(A_{11}, ..., A_{nn}) = \sum_S ||x_i - x_j||_A^2 - log(\sum_D ||x_i - x_j||_A)$$

Let

$$\begin{aligned} &x_{ij} = [(x_{i1} - x_{j1})^2, ..., (x_{id} - x_{jd})^2)]^T \\ &A = [A_{11}, ..., A_{dd}] \end{aligned}$$

Here

$$\begin{aligned} &||x_i - x_j||_A^2 = (x_i - x_j)^T A (x_i - x_j) = x_{ij}^T A \\ &g(A) = \sum_S x_{ij}^T A - log(\sum_D (x_{ij}^T A)^{\frac{1}{2}}) \end{aligned}$$

Then

$$\begin{aligned} g'(A) &= \sum_S x_{ij}^T A - log(\sum_D (x_{ij}^T A)^{\frac{1}{2}}) \\ &= \sum_S x_{ij} - \frac{1}{\sum_D (x_{ij}^T A)^{\frac{1}{2}}} \sum_D \frac{1}{2} (x_{ij}^T A)^{-\frac{1}{2}} x_{ij} \end{aligned}$$

$g'(A)$ is a vector with the same size as A. In this case, we want to use Newton-Raphson method to get the parameter A, so we need to derivate the Hessian Matrix ($g''(A)$) to be used in the update rule. Here we give the derivation as follows.

$$
\begin{aligned}
&g''(A) = \partial g'(A)^T / \partial A \\
&= \partial[-[\sum_D (x_{ij}^T A)^{\frac{1}{2}}]^{-1} \sum_D \frac{1}{2}(x_{ij}^T A)^{-\frac{1}{2}} x_{ij}^T] / \partial A \\
&= [\sum_D (x_{ij}^T A)^{\frac{1}{2}}]]^{-2} \sum_D \frac{1}{2}(x_{ij}^T A)^{\frac{1}{2}} x_{ij} \sum_D \frac{1}{2}(x_{ij}^T A)^{-\frac{1}{2}} x_{ij}^T - \\
&[\sum_D (x_{ij}^T A)^{\frac{1}{2}}]^{-1} \sum_D -\frac{1}{4}(x_{ij}^T A)^{-\frac{3}{2}} x_{ij} x_{ij}^T
\end{aligned}
$$