A Quadratic Program Formulation for Spectral Transformation

Xiaojin Zhu

April 10, 2005

An *order constrained* semi-supervised kernel K is the solution to the following convex optimization problem:

$$\begin{aligned} \max_K \quad & \hat{A}(K_{tr}, T) && (1)\\ \text{subject to} \quad & K = \textstyle\sum_{i=1}^n \mu_i K_i && (2)\\ & \mu_i \geq 0 && (3)\\ & \operatorname{trace}(K) = 1 && (4)\\ & \mu_i \geq \mu_{i+1}, \quad i = 1 \cdots n-1 && (5) \end{aligned}$$

where T is the training target matrix, $K_i = \phi_i \phi_i^\top$ and ϕ_i's are the eigenvectors of the graph Laplacian.

This formulation is an extension to the original kernel alignment with the addition of order constraints, and with special components K_i's from the graph Laplacian. Since $\mu_i \geq 0$ and K_i's are outer products, K will automatically be positive semi-definite and hence a valid kernel matrix. It is important to notice the order constraints are convex, and as such it is a convex optimization problem. Kernel alignment is invariant to scales. That is, if we scale μ by an arbitrary positive constant a, the kernel $K' = \sum_{i=1}^n a\mu_i K_i$ will have the same alignment as K. The trace constraint is one way to fix the scale invariance of kernel alignment.

An alternative way to fix scale invariance without the trace constraint is to note that the objective is equivalent to

$$\frac{\langle K_{tr}, T \rangle_F}{\sqrt{\langle K_{tr}, K_{tr} \rangle_F}} \tag{6}$$

where a scaling constant a would appear in the numerator and denominator at the same time and cancel out. Therefore we can fix the numerator to 1 and minimize the denominator, or fix the denominator and maximize the numerator. The former has a quadratic objective function and a linear constraint, which is a quadratic program (QP). The latter has a linear objective function but a quadratic constraint, which is a special case of the quadratically constrained quadratic program (QCQP). In both cases the solution is equivalent (up to scaling) to the original problem with the trace constraint.

We prefer the QP problem because it is simpler. The problem can be rewritten as

$$\begin{aligned}
\min_\mu \qquad & \sqrt{\langle K_{tr}, K_{tr} \rangle_F} && (7)\\
\text{subject to} \qquad & K = \textstyle\sum_{i=1}^{n} \mu_i K_i && (8)\\
& \langle K_{tr}, T \rangle_F = 1 && (9)\\
& \mu_i \geq 0 && (10)\\
& \mu_i \geq \mu_{i+1}, \quad i = 1 \cdots n-1 && (11)
\end{aligned}$$

Let $vec(A)$ be the column vectorization of a matrix A. Defining

$$M = [vec(K_{1,tr}) \cdots vec(K_{m,tr})] \tag{12}$$

It is not hard to show that the problem can then be expressed as

$$\begin{aligned}
\min_\mu \qquad & ||M\mu|| && (13)\\
\text{subject to} \qquad & vec(T)^\top M\mu = 1 && (14)\\
& \mu_i \geq 0 && (15)\\
& \mu_i \geq \mu_{i+1}, \quad i = 1 \cdots n-1 && (16)
\end{aligned}$$

Finally minimizing the norm is equivalent to minimizing the squared norm

$$\begin{aligned}
\min_\mu \qquad & \mu^\top M^\top M\mu && (17)\\
\text{subject to} \qquad & vec(T)^\top M\mu = 1 && (18)\\
& \mu_i \geq 0 && (19)\\
& \mu_i \geq \mu_{i+1}, \quad i = 1 \cdots n-1 && (20)
\end{aligned}$$

The objective function is quadratic in μ, and the constrains are linear, making it a quadratic program (QP).