Optimal Kernel Hashing with Multiple Features

supplemental material: Proof of Equation (7) and (8)

APPENDIX

Given a set of N training examples with M visual features, the m-th feature (d_m dimension) of n-th sample can be represented as $X_n^{(m)} \in R^{d_m \times 1}$. Then $X^{(m)} = [X_1^{(m)}, X_2^{(m)}, \ldots, X_N^{(m)}] \in R^{d_m \times N}$ is the m-th feature matrix of all training data.

In order to obtain compact hash codes, we give a formulation similar to Spectral Hashing. The hash codes are learned to preserve similarity between data points, and meanwhile satisfying balance and independent constrains. Unlike the previous methods, the hash codes are explicitly related to the kernels corresponding to each visual feature:

$$\begin{aligned} \min_{W,b,\mu} \quad & \frac{1}{2}\sum_{i,j=1}^{N} S_{ij}\|Y_i - Y_j\|^2 + \lambda\|V\|_F^2 \\ s.t. \quad & Y_i \in \{-1,1\}^P \\ & \sum_{i=1}^{N} Y_i = 0 \\ & \frac{1}{N}\sum_{i=1}^{N} Y_i Y_i^T = I \\ & \mathbf{1}^T\mu = 1,\ \mu \succeq 0 \end{aligned} \tag{1}$$

where

$$\varphi(X_i) = [\mu_1^{\frac{1}{2}}\varphi_1^T(X_i^{(1)}), \ldots, \mu_M^{\frac{1}{2}}\varphi_M^T(X_i^{(M)})]^T \tag{2}$$

$$V_p = \sum_{l=1}^{L} W_{lp}\varphi(Z_l),\ p = 1, \ldots, P \tag{3}$$

and

$$Y_{pi} = h_p(X_i) = sign(V_p^T\varphi(X_i) + b_p),\ i = 1, \ldots, N \tag{4}$$

Similar to spectral hashing, the discrete constrains of $Y_i \in \{-1, 1\}$ are relaxed and then $Y_{pi} = h_p(X_i) = V_p^T\varphi(X_i) + b_p$. For each feature, define its kernel as $K^{(m)}$ corresponding to its embedding function $\varphi_m(\cdot)$, which means that $K_{ij}^{(m)} = \varphi_m(X_i^{(m)})^T\varphi_m(X_j^{(m)})$. With the above definition,

$$\begin{aligned} K_{ij} &= \varphi(X_i)^T\varphi(X_j) \\ &= \sum_{m=1}^{M} (\mu_m^{\frac{1}{2}}\varphi_m(X_i^{(m)}))^T(\mu_m^{\frac{1}{2}}\varphi_m(X_j^{(m)})) \\ &= \sum_{m=1}^{M} \mu_m K_{ij}^{(m)} \end{aligned} \tag{5}$$

Therefore, $K = \sum_{m=1}^{M} \mu_m K^{(m)}$, and Y_i can be reformulated in kernel form:

$$Y_i = V^T\varphi(X_i) + b = W^T K_i + b \tag{6}$$

where K_i is the i-th column of K, and $b = [b_1, b_2, \ldots, b_P]$.

A. DERIVATION OF PROBLEM (7)

From constraint

$$\sum_{i}^{N} Y_i = \sum_{i}^{N} W^T K_i + b = 0 \tag{7}$$

we get

$$b = -\frac{1}{N}W^T K_{L\times N}\mathbf{1} \tag{8}$$

Then substitute b into Y_i, and rewrite the objective as:

$$\begin{aligned} & \frac{1}{2}\sum_{i,j=1}^{N} S_{ij}\|Y_i - Y_j\|^2 + \lambda\|V\|_F^2 \\ &= \mathrm{tr}(W^T K_{L\times N}(\Delta - S)K_{L\times N}^T W) + \lambda tr(W^T K_{L\times L}W) \\ &= \mathrm{tr}(W^T(K_{L\times N}(\Delta - S)K_{L\times N}^T + \lambda K_{L\times L})W) \\ &= \mathrm{tr}(W^T C W) \end{aligned} \tag{9}$$

where $C = K_{L\times N}(\Delta - S)K_{L\times N}^T + \lambda K_{L\times L}$ and $\Delta = diag(S\mathbf{1})$. In the above derivation, we use the fact that

$$(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)(\Delta - S)(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)^T = \Delta - S. \tag{10}$$

For another constraint $\frac{1}{N}\sum_{i=1}^{N} Y_i Y_i^T = I$, substitute Y_i

MM '11 El Paso, Texas USA

and b in into it and we get

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} Y_i Y_i^T &= \frac{1}{N}\sum_{i=1}^{N}(W^T K_i + b)(W^T K_i + b)^T \\ &= \frac{1}{N} W^T K_{L\times N}(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)^T K_{L\times N}^T W \\ &= \frac{1}{N} W^T K_{L\times N}(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T) K_{L\times N}^T W\end{aligned} \tag{11}$$

Here we use the fact that

$$(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)^T = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T. \tag{12}$$

Therefore the constrain turns to be $W^T G W = I$, where $G = \frac{1}{N} K_{L\times N}(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T) K_{L\times N}^T$.

In summary, given the fixed μ, the optimal W and b can be obtained by solving the following problem:

$$\begin{aligned}\min_W \quad & \mathrm{tr}(W^T C W) \\ s.t. \quad & W^T G W = I\end{aligned} \tag{13}$$

where

$$C = K_{L\times N}(\Delta - S) K_{L\times N}^T + \lambda K_{L\times L}$$

$$G = \frac{1}{N} K_{L\times N}(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T) K_{L\times N}^T.$$

Here $\Delta = diag(S\mathbf{1})$ and $b = -\frac{1}{N} W^T K_{L\times N}\mathbf{1}$. Such problem can be optimized efficiently by eigen-decomposition.

B. DERIVATION OF PROBLEM (8)

Given W and b, the objective in Equation 9 can be written as

$$\begin{aligned}&\frac{1}{2}\sum_{i,j=1}^{N} S_{ij}\|Y_i - Y_j\|^2 + \lambda\|V\|_F^2 \\ &= \mathrm{tr}(W^T(\sum_{m=1}^{M}\mu_m K_{L\times N}^{(m)})(\Delta - S)(\sum_{m=1}^{M}\mu_m K_{L\times N}^{(m)})^T W) \\ &\quad + \lambda\mathrm{tr}(W^T((\sum_{m=1}^{M}\mu_m K_{L\times L}^{(m)}))W) \\ &= \mathrm{tr}(\sum_{i,j=1}^{M}\mu_i\mu_j W^T K_{L\times N}^{(i)}(\Delta - S) K_{L\times N}^{(j)} W) \\ &\quad + \lambda\mathrm{tr}(\sum_{i=1}^{M}\mu_i W^T K_{L\times L}^{(i)} W) \\ &= \frac{1}{2}\mu^T E\mu + h^T\mu\end{aligned} \tag{14}$$

where $E_{ij} = 2tr(W^T K_{L\times N}^{(i)}(\Delta - S){K_{L\times N}^{(j)}}^T W)$ and $h_i = tr(W^T K_{L\times L}^{(i)} W)$.

Therefore the the optimization with respect to μ can be formulated as a quadratic programming problem as follows:

$$\begin{aligned}\min_\mu \quad & \frac{1}{2}\mu^T E\mu + h^T\mu \\ s.t. \quad & \mathbf{1}^T\mu = 1, \mu \succeq 0.\end{aligned} \tag{15}$$

where,

$$E_{ij} = 2\mathrm{tr}(W^T K_{L\times N}^{(i)}(\Delta - S){K_{L\times N}^{(j)}}^T W), i, j = 1, \ldots, M$$

$$h_i = \lambda\mathrm{tr}(W^T K_{L\times L}^{(i)} W), i = 1, \ldots, M$$