5 HSIC to MLE

Assume that we are maximizing HSIC in the IDS space. The formulation becomes

$$\max_{W} \quad \mathrm{Tr}(\Phi W W^T \Phi H K_y H) \tag{18}$$

$$\max_{W} \quad \mathrm{Tr}(W^T \Phi^T H K_y H \Phi W). \tag{19}$$

Without a loss of generality, for now assume that $W \in \mathbb{R}^{d \times q}$ and $q = 1$. Under our *Kernel Chain* formulation, W is simply a vector in RKHS that can represent a function. Here we will let $W = f$ to emphasize its role as a function. We also not the reproducing property of the function in RKHS where

$$f(x) = \langle f, \phi(x) \rangle. \tag{20}$$

Next, we denote $\hat{\Phi} = H\Phi$ as feature maps that have already been centered. The objective becomes

$$\max_{f} \mathrm{Tr}(f^T \hat{\Phi} Y Y^T \hat{\Phi} f) \tag{21}$$

$$\max_{f} \mathrm{Tr}(f^T \begin{bmatrix} \hat{\phi}(x_1) & ... & \hat{\phi}(x_n) \end{bmatrix} Y Y^T \hat{\Phi} f). \tag{22}$$

Since each column of Y is an indicator vector indicating if a vector belongs to a class, we get

$$\max_{f} \quad \mathrm{Tr}(f^T \begin{bmatrix} \sum_{i \in \mathcal{S}^1} \hat{\phi}(x_i) & ... & \sum_{i \in \mathcal{S}^c} \hat{\phi}(x_c) \end{bmatrix} Y^T \hat{\Phi} f) \tag{23}$$

$$\max_{f} \quad \mathrm{Tr}(\begin{bmatrix} \sum_{i \in \mathcal{S}^1} f(x_i) & ... & \sum_{i \in \mathcal{S}^c} f(x_c) \end{bmatrix} Y^T \hat{\Phi} f) \tag{24}$$

$$\max_{f} \quad \mathrm{Tr}(\begin{bmatrix} \sum_{i \in \mathcal{S}^1} f(x_i) & ... & \sum_{i \in \mathcal{S}^c} f(x_c) \end{bmatrix} \begin{bmatrix} \sum_{i \in \mathcal{S}^1} f(x_i) & ... & \sum_{i \in \mathcal{S}^c} f(x_c) \end{bmatrix}^T) \tag{25}$$

$$\max_{f} \quad \left[\sum_{i \in \mathcal{S}^1} f(x_i) \right]^2 + ... + \left[\sum_{i \in \mathcal{S}^c} f(x_i) \right]^2 \tag{26}$$

$$\max_{f} \quad \exp \left\{ \left[\sum_{i \in \mathcal{S}^1} f(x_i) \right]^2 + ... + \left[\sum_{i \in \mathcal{S}^c} f(x_i) \right]^2 \right\} \tag{27}$$

$$\max_{f} \quad \exp \left\{ \left[\sum_{i \in \mathcal{S}^1} f(x_i) \right]^2 \right\} \quad ... \quad \exp \left\{ \left[\sum_{i \in \mathcal{S}^c} f(x_i) \right]^2 \right\} \tag{28}$$

The final objective is the maximum likelihood for an exponential family distribution.