Homework 3 Solution

Problem 1. *Given two vectors of real numbers* $X = \{x_1, \dots, x_n\} \in \mathbb{R}^n$ *and* $Y = \{y_1, \dots, y_n\} \in \mathrm{R}^n$*, the covariance between* X *and* Y *is defined as*

$$cov_n(X, Y) = \mathrm{E}_n(XY) - \mathrm{E}_n(X)\mathrm{E}_n(Y)$$

where $\mathrm{E}_n(U) = (\sum_{i=1}^n u_i)/n$. *The covariance is useful to detect linear relationships between* X *and* Y*. In order to extend this measure to potential nonlinear relationships between* X *and* Y*, we consider the following criterion:*

$$C_n^K(X, Y) = \max_{f,g \in \mathcal{B}_K} cov_n(f(X), g(Y))$$

where K *is a positive definite kernel on* $\mathbb{R}$, $\mathcal{B}_K$ *is the unit ball of the RKHS of* K*, and* $f(U) = (f(u_1), \dots, f(u_n))$ *for a vector* $U = (u_1, \dots, u_n)$.

1. *Express simply* $C_n^K(X, Y)$ *for the linear kernel* $K(a, b) = ab$.
2. *For a general kernel* K*, express* $C_n^K(X, Y)$ *in terms of the Gram matrices of* X *and* Y*.*

Let's begin with the general case and then check for the linear kernel. Assuming that our data is centered, we can write $C_n^K(X, Y)$ as

$$\min_{f,g \in \mathcal{B}_K} - \sum_{i=1}^{n} f(x_i)g(y_i)$$

note that the sum can be written in the matrix form $\alpha^\top K_x K_y \beta$, thus the Lagrangian can be written as:

$$\mathcal{L}(\alpha, \beta, \lambda_x, \lambda_y) = -\alpha^\top K_x K_y \beta + \lambda_x(\alpha^\top K_x \alpha - 1) + \lambda_y(\beta^\top K_y \beta - 1)$$

setting the partial derivatives with respect to α and β yields:

$$\begin{aligned} 2\lambda_x \alpha - K_y \beta &= 0 \\ 2\lambda_y \beta - K_x \alpha &= 0 \end{aligned} \tag{1}$$

multiply the first equation by $\alpha^\top K_x$ and the second one by $\beta^\top K_y$

$$2\lambda_x \alpha^\top K_x \alpha - \alpha^\top K_x K_y \beta = 0$$

$$2\lambda_y \beta^\top K_y \beta - \beta^\top K_y K_x \alpha = 0$$

subtracting the two, we get:

$$\lambda_x \alpha^\top K_x \alpha = \lambda_y \beta^\top K_y \beta = \lambda_x = \lambda_y = \lambda$$

we can thus rewrite (1) in matrix from:

$$M \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda' \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

where $\lambda' = 2\lambda$ and $M = \begin{bmatrix} 0 & K_y \\ K_x & 0 \end{bmatrix}$. Therefore, the answer would be given by the eigen pairs of M.

$$det(M - \lambda' I) = 0 \Rightarrow det({\lambda'}^2 I - K_y K_x) = 0$$

so $\lambda' = \sqrt{\lambda''}$ where λ'' is an eigenvalue of $K_y K_x$. note that λ'' is non-negative since the matrix is positive semi-definite. this means that the answer is given by the eigenvectors of M with eigenvalue $\sqrt{\lambda''}$.

Now we return to the first question. Note that $K_x = XX^\top$ for the linear kernel. so λ'' is the eigenvalue of $YY^\top XX^\top$.
note that $Y^\top X$ is a scaler, denoting it with r, the matrix can be written as $rYX^\top$.
note that $YX^\top$ is a rank-one matrix, which means that it has at most one non-zero eigenvalue, which is given by the trace. Further, the trace of $YX^\top$ is r, which implies that $\lambda'' = r^2$
This means that the answer is given by the eigenvectors of M with eigenvalue r. □