1. Denote \( \sigma(t) = \frac{1}{1 + e^{-t}} \). Verify that the ERM corresponding to the logistic loss over the functions of the form \( f(x) = w^\top \varphi(x) \) can be written as

\[
\min_w \sum_{i=1}^n - \log \sigma(y_i w^\top \varphi(x_i)) + \lambda \|w\|^2
\]

and is a convex optimisation problem in \( w \). By the representer theorem, we can write

\[ w = \sum_{i=1}^n \alpha_i \varphi(x_i) \]

Show that the criterion in (1) is also convex in the dual coefficients \( \alpha \in \mathbb{R}^n \).

[Hint: \( \sigma'(t) = \sigma(t) \sigma(-t) \)]

2. Let \( k_1 \) and \( k_2 \) be positive definite kernels on \( \mathbb{R}^p \). Verify that the following are also valid kernels.

[Hint: it suffices to identify the corresponding feature.]

(a) \( x^\top x' \),

(b) \( ck_1(x, x') \), for \( c \geq 0 \),

(c) \( f(x)k_1(x, x')f(x') \) for any function \( f : \mathbb{R}^p \to \mathbb{R} \),

(d) \( k_1(x, x') + k_2(x, x') \),

(e) \( k_1(x, x')k_2(x, x') \),

(f) \( \exp(k_1(x, x')) \),

(g) \( \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|^2\right) \).

3. Assume that kernel \( k \) is not strictly positive definite, but that there exist \( \{a_i\}_{i=1}^n \) and \( \{x_i\}_{i=1}^n \), such that

\[
\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(x_i, x_j) = 0.
\]

Show that then

\[ f(x) = \sum_{i=1}^n a_i k(x_i, x) = 0 \quad \forall x \in \mathcal{X}. \]

Hence conclude that the RKHS functions of the form \( f(x) = \sum_{i=1}^n a_i k(x_i, x) \) have zero norm if and only if they are identically equal to zero. [Hint: assume contrary for some \( x = x_{n+1} \) and consider \( \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a_i a_j k(x_i, x_j) \)]

4. (One-Class SVM) A Gaussian RBF kernel on \( \mathcal{X} = \mathbb{R}^p \) is given by

\[
k(x, x') = \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|^2\right). \tag{2}
\]

(i) What is \( k(x, x) \) for this kernel? What can you conclude about the norm of the features \( \varphi(x) \) of \( x \)? What values can the angles between \( \varphi(x) \) and \( \varphi(x') \) take? Sketch the set \( \{\varphi(x) : x \in \mathcal{X}\} \) as if the features lived in a 2D space.
(ii) Let \( \{x_i\}_{i=1}^n \) be a set of points in \( \mathcal{X} = \mathbb{R}^p \) (no labels are given). The one-class Support Vector Machine (SVM) is a method for outlier detection which in its primal form is defined as

\[
\min_{w, \xi, \rho} \frac{1}{2} \|w\|^2 + \frac{1}{\nu n} \sum_{i=1}^n \xi_i - \rho, \quad \text{subject to } \langle w, \varphi(x_i) \rangle \geq \rho - \xi_i, \; \xi_i \geq 0,
\]

where \( \nu \) is a given SVM parameter, features \( \varphi(x) \) correspond to the RBF kernel in \( \mathcal{F} \), and \( \xi_i \)'s are the non-negative slack variables. The fitted hyperplane \( \langle w, \varphi(x) \rangle - \rho \) in the feature space separates the majority of points from the origin (while pushing away from the origin as much as possible) and is used to determine “atypical” \( x \)-instances.

Using the 2D intuition from (i), sketch the corresponding hyperplane in the feature space and annotate with \( \rho \), \( w \) and a non-zero slack \( \xi_j \) for an “outlier” \( x_j \). Would it make sense to use the one-class SVM with a linear kernel?

(iii) Write the dual form of the one-class SVM, using Lagrangian duality.

[Hint: setting to zero the derivative of the Lagrangian with respect to \( w \) should give \( w = \sum_{i=1}^n \alpha_i \varphi(x_i) \), where \( \alpha_i \geq 0 \) are the Lagrange multipliers of the constraints \( \langle w, \varphi(x_i) \rangle \geq \rho - \xi_i \)].

5. Under the assumption that your data are centred, show that you can compute the \( n \times n \) Gram matrix \( K \) such that \( K_{ij} = x_i^\top x_j \) using the dissimilarity matrix \( D \) where \( D_{ij} = \|x_i - x_j\|_2 \).

6. Show that

\[
\text{MMD}_k(P, Q) = \sup_{f \in \mathcal{H}_k : \|f\|_{\mathcal{H}_k} \leq 1} \left| \mathbb{E}_{X \sim P} f(X) - \mathbb{E}_{Y \sim Q} f(Y) \right|.
\]

7. Let \( L \) be an unnormalized Laplacian matrix of a graph with \( C \) connected components. Verify that

(a) Column vector \( \mathbf{1} \) is the eigenvector of \( L \) with eigenvalue 0.

(b) \( L \) is positive semi-definite.

(c) \( v \) is an eigenvector of \( L \) corresponding to 0-eigenvalue if and only if \( v \in \text{span}\{e_1, \ldots, e_C\} \), where

\[
e_{ci} = \begin{cases} 
1, & \text{vertex } i \text{ belongs to the connected component } c, \\
0, & \text{otherwise}.
\end{cases}
\]

8. Verify that for a given partition \( C_1, C_2, \ldots, C_K \) and column vectors \( h_k \in \mathbb{R}^n \) defined as \( h_k = \frac{1}{\sqrt{|C_k|}} \mathbf{1}_{\{i \in C_k\}} \), we have

\[
\text{ratio-cut} (C_1, \ldots, C_K) = \sum_{k=1}^K h_k^\top L h_k.
\]