1. Let V be a finit dimensional vector space equipped with an inner product $< \cdot,\cdot >_V$. The vector space of linear maps $V \to V$ is isomorphic to the bilinear functionals $V \times V \to \mathbb{R}$. The subspace of self-adjoint maps is isomorphic to the symmetric biliniear functionals. Therefore an inner product on V must have the form $(x,y) \to < x, My >_V$ for some selft adjoint linear map $M : V \to V$. Denote this map by $< \cdot,\cdot >_M$. Prove that M is positive definite iff $< \cdot,\cdot >_M$ is an inner product.

2. Suppose $\mathcal{X}$ is a finit set of size n equipped with the counting measure. Consider the set of functions $\mathcal{X} \to \mathbb{R}$. This is an n -dimensional Hilbert space V. We can associate a symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with a self-adjoin linear map $K : V \to V$,

$$(Ku)_i = \sum_{j=1}^{n} k(x_i, k_j) u_j$$

.

We say k is positive definite iff corresponding transformation k is positive definite. Now prove that a symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite iff k is a kernel.

3. **Mercer's Theorem** If Gram matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ is positive definite, each element can be represented as:

$$\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

where $\phi : \chi \to D$ is the feature map to higher dimensional space. This can be proved by diagonalizing the Gram matrix as:

$$K = U^H \Lambda U$$

where Λ is the diagonal matrix whose elements are the eigen values. Since it is diagonal, it can be directly factored, and K can be written as :

$$K = (\Lambda^{\frac{1}{2}} U)^H \ \ \Lambda^{\frac{1}{2}} U$$
$$\kappa(x_i, x_j) = (\Lambda^{\frac{1}{2}} U_{i,:})^T \ \ \Lambda^{\frac{1}{2}} U_{:,j}$$
$$\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

   (a) Show that the Gram Matrix **K** defined by **inner product** : $\kappa(x_i, x_j) =< x_i, x_j >$ is Positive Definite.

   (b) Gaussian Kernel is defined by: $\kappa_{i,j} = e^{-\frac{||x_i - xj||^2}{2\sigma^2}}$. Express this is as inner product of feature maps in the Hilbert Space.

4. **Cauchy-Schwarz for RKHS.** Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a PDS kernel. For any $x \in \mathcal{X}$, define $\Phi_x : \mathcal{X} \to \mathbb{R}$ for all $x' \in \mathcal{X}$ as follows:

$$\Phi_x(x') = K(x, x').$$

We define $\mathbb{H}_0$ as the set of finite linear combinations of such functions Φ_x:

$$\mathbb{H}_0 = \Big\{ \sum_{i \in I} a_i \Phi_{x_i} : a_i \in \mathbb{R}, x_i \in \mathcal{X}, |I| < \infty \Big\}.$$

Now, we introduce the operation $\langle .,. \rangle$ on $\mathbb{H}_0 \times \mathbb{H}_0$ defined for all $f, g \in \mathbb{H}_0$ with $f = \sum_{i \in I} a_i \Phi_{x_i}$ and $g = \sum_{j \in J} b_j \Phi_{x_j}$ by

$$\langle f, g \rangle = \sum_{i \in I, j \in J} a_i b_j K(x_i, x_j).$$

Show that for any $f \in \mathbb{H}_0$ and any $x \in \mathcal{X}$,

$$\langle f, \Phi_x \rangle^2 \leqslant \langle f, f \rangle \langle \Phi_x, \Phi_x \rangle.$$

5. Let k_1, k_2 be kernels over $\mathbb{R}^n \times \mathbb{R}^n$, and k_3 be a kernel over $\mathbb{R}^d \times \mathbb{R}^d$. Let $a \in \mathbb{R}_+$ be a positive real number, $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function, $\phi : \mathbb{R}^n \to \mathbb{R}^d$ be a function mapping from $\mathbb{R}^n$ to $\mathbb{R}^d$, and let p be a polynomial over $\mathbb{R}_+$ with positive coefficients. For each of the functions k below, state whether it is necessarily a kernel. If you think it is, prove it; if you think it isn't, give a counter-example.
   (a) $k(x,y) = k_1(x,y) + k_2(x,y)$
   (b) $k(x,y) = k_1(x,y) - k_2(x,y)$
   (c) $k(x,y) = a\, k_1(x,y)$
   (d) $k(x,y) = -a\, k_1(x,y)$
   (e) $k(x,y) = k_1(x,y)\, k_2(x,y)$
   (f) $k(x,y) = f(x)\, f(y)$
   (g) $k(x,y) = k_3\Big(\phi(x), \phi(y)\Big)$
   (h) $k(x,y) = p\Big(k_1(x,y)\Big)$

6. **Axis-aligned rectangles.** Consider, $\mathcal{X} = \mathbb{R}^2$ and the concept class $C \subseteq \{0,1\}^{\mathcal{X}}$ is the set of all axis-aligned rectangles in $\mathbb{R}^2$. Thus, each concept c is the set of points inside a particular axis-aligned rectangle. We note that an unlabeled sample $x \in \mathcal{X}^m$ consist of feature vectors $x_i = (x_{i1}, x_{i2}) \in \mathbb{R}^2$ for all $i \in [m]$. A concept $c \in C$ is defined by four real numbers l, r, b, t, such that

$$c(x_i) = \mathbb{1}_{[l,r]}(x_{i1})\mathbb{1}_{[b,t]}(x_{i2}).$$

   (a) Given a labeled sample $z \in (\mathcal{X} \times \{0,1\})^m$, let $S = \{i \in [m] : y_i = 1\}$ be the indices of the point inside the rectangle. Then, the data driven concept $\hat{c} : \mathcal{X} \to \{0,1\}$ is proposed in terms of parameters $(\hat{l}, \hat{r}, \hat{b}, \hat{t})$ defined as

$$\hat{l} = \inf\{x_{i1} : i \in S\}, \quad \hat{r} = \sup\{x_{i1} : i \in S\}, \quad \hat{b} = \inf\{x_{i2} : i \in S\}, \quad \hat{t} = \sup\{x_{i2} : i \in S\}.$$

   Find the empirical loss defined as $\hat{R}(\hat{c}) \triangleq \frac{1}{m}\sum_{i=1}^m \mathbb{1}_{\{y_i \neq \hat{c}(x_i)\}}$. **(2)**

   (b) Let $D : \mathcal{B}(\mathcal{X}) \to [0,1]$ be an unknown distribution for a point location in $\mathbb{R}^2$, and an unlabeled sample is drawn *i.i.d.* . The (l,r,b,t) be the parameters for the true concept $c \in C$, such that the probability of a point $X_i \in \mathcal{X}$ drawn according to distribution D is labeled 1 is given by

$$p(l,r,b,t) \triangleq \mathbb{E}_{x_i}\mathbb{1}_{[l,r]}(x_{i1})\mathbb{1}_{[b,t]}(x_{i2}).$$

   Find the generalization risk written as $R(\hat{c}) = \mathbb{E}_{x_j}\left\{c(x_j) \neq \hat{c}_z(x_j)\right\}$ in term of the given labeled sample $z \in (\mathcal{X} \times \{0,1\})^m$. **(2)**

   (c) Can you show that $P_x\{R(\hat{c}) > \varepsilon\} \leqslant 4(1-\varepsilon/4)^m$? **(1)**

7. **XOR Problem** Consider a XOR problem using SVMs, now use kernel $K(\underline{x}, \underline{x}_i) = (1 + \underline{x}^T \underline{x}_i)^2$ to solve the problem. Give the optimum Hyperplane equation.

8. Consider the data set `https://archive.ics.uci.edu/ml/datasets/iris`. We are intrested in consutructing a linear classifier for this data based on SVM.
   Input:
   1. Training set (X, y)
   2. Parameter C
   3. Kernel type (linear/polynomial/Gaussian)
   4. Kernel parameter

   Kernels:

1. Linear
2. Polynomial kernel $k(x,y) = (x^T y + 1)^d$, Kernel parameter = d
3. Gaussian kernal $k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2g^2}\right)$, Kernel parameter = g

Output:

Program should return as output an SVM model and error as you did in Assignment 1. You can use your first homework as a starting point.