Assignment Five
ECE 4200/5420

- Provide credit to **any sources** other than the course staff that helped you solve the problems. This includes **all students** you talked to regarding the problems.
- You can look up definitions/basics online (e.g., wikipedia, stack-exchange, etc).
- **The due date is 10/18/2021, 23.59.59 ET**.
- Submission rules are the same as previous assignments.
- **Please write your net-id on top of every page. It helps with grading.**

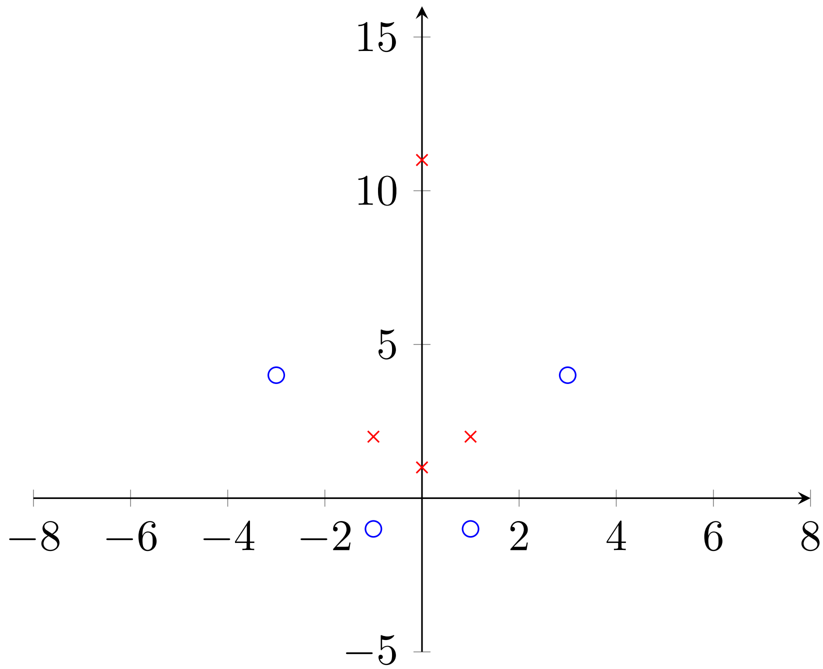
Problem 1. (15 points). SVM's obtain *non-linear* decision boundaries by mapping the feature vectors $X \in \mathbb{R}^d$ to a possibly high dimensional space via a function $\phi : \mathbb{R}^d \to \mathbb{R}^m$, and then finding a linear decision boundary in the new space.

We also saw that to implement SVM, it suffices to know the kernel function $K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j)$, without even explicitly specifying the function ϕ.

Recall **Mercer's theorem**. K is a kernel function if and only if for any n vectors, $X_1, \ldots, X_n \in \mathbb{R}^d$, and **any** real numbers $c_1, \ldots, c_n$, $\sum_{i=1}^n \sum_{j=1}^n c_i c_j K(X_i, X_j) \geq 0$.

1. Prove the following half of Mercer's theorem (which we showed in class). If K is a kernel then $\sum_{i=1}^n \sum_{j=1}^n c_i c_j K(X_i, X_j) \geq 0$.
2. Let $d = 1$, and $x, y \in \mathbb{R}$. Is the function $K(x, y) = x + y$ a kernel?
3. Let $d = 1$, and $x, y \in \mathbb{R}$. Is $K(x, y) = xy + 1$ a kernel?
4. Suppose $d = 2$, namely the original features are of the form $X = [X^1, X^2]$. Show that $K(X, Y) = (1 + \bar{X} \cdot \bar{Y})^2$ is a kernel function. This is called as **quadratic kernel**.

   (**Hint**: Find a $\phi : \mathbb{R}^2 \to \mathbb{R}^m$ (for some m) such that $\phi(X) \cdot \phi(Y) = (1 + X \cdot Y)^2$).
5. Consider the training examples $\langle [0, 1], 1\rangle$, $\langle [1, 2], 1\rangle$, $\langle [-1, 2], 1\rangle$, $\langle [0, 11], 1\rangle$, $\langle [3, 4], -1\rangle$, $\langle [-3, 4], -1\rangle$, $\langle [1 - 1], -1\rangle$, $\langle [-1, -1], -1\rangle$. We have plotted the data points below.
   - Is the data **linearly classifiable** in the original 2-d space? If yes, please come up with *any* linear decision boundary that separates the data. If no, please explain why.
   - Is the data linearly classifiable in the feature space corresponding to the quadratic kernel. If yes, please come up with *any* linear decision boundary that separates the data. If no, please explain why.



Problem 2. (10 points). The Gaussian kernel (also called Radial Basis Function kernel (RBF)) is:

$$K(X, Y) = \exp\left(-\frac{\|X - Y\|_2^2}{2\sigma^2}\right),$$

where X, Y are feature vectors in d dimensions. Suppose $d = 1$, and $2\sigma^2 = 1$.

1. Design a function $\phi : \mathbb{R} \to \mathbb{R}^m$ that corresponds to Gaussian kernel for $d = 1$, and $2\sigma^2 = 1$.

   Hint: Use Taylor series expansion for the exponential function.

2. What is the value of m you end up with?

Problem 3. (10 points). Let $f, h_i, 1 \leq i \leq n$ be functions from $\mathbb{R}^m \to \mathbb{R}$ (real-valued functions) and let $\boldsymbol{\alpha} \in \mathbb{R}^n$. Let $L(z, \boldsymbol{\alpha}) = f(z) + \sum_{i=1}^{n} \alpha_i h_i(z)$. In this problem, we will prove that the following two optimization problems are equivalent.

$$\begin{aligned} &\min_{\boldsymbol{z}} f(\boldsymbol{z}) \\ &\text{s.t. } h_i(\boldsymbol{z}) \leq 0,\ i = 1, \ldots, n. \end{aligned} \quad (1) \qquad\qquad \min_{\boldsymbol{z}} \max_{\boldsymbol{\alpha} \geq \boldsymbol{0}} L(\boldsymbol{z}, \boldsymbol{\alpha}) \quad (2)$$

Let $(\boldsymbol{z}^*, \boldsymbol{\alpha}^*)$ be the solution of (2) and let $\boldsymbol{z}_p^*$ be the solution of (1).

1. Prove that:

$$L(\boldsymbol{z}^*, \boldsymbol{\alpha}^*) = f(\boldsymbol{z}_p^*).$$

2. **(complementary slackness).** Let $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_n^*)$. Show that for all $1 \leq i \leq n$:

$$\boldsymbol{\alpha}_i^* \cdot h_i(\boldsymbol{z}^*) = 0.$$

Hint: For $\boldsymbol{z} \in \mathbb{R}^m$, let $\boldsymbol{\alpha}_{\boldsymbol{z}} = \arg\max_{\boldsymbol{\alpha} \geq \boldsymbol{0}} L(\boldsymbol{z}, \boldsymbol{\alpha})$, which is a function of $\boldsymbol{z}$. Then note that (2) is the same as $\min_{\boldsymbol{z}} L(\boldsymbol{z}, \boldsymbol{\alpha}_{\boldsymbol{z}})$ Use the fact that for any $\boldsymbol{z}$, $\boldsymbol{\alpha} \geq \boldsymbol{0}$, $L(\boldsymbol{z}^*, \boldsymbol{\alpha}^*) \geq L(\boldsymbol{z}^*, \boldsymbol{\alpha})$ and $L(\boldsymbol{z}^*, \boldsymbol{\alpha}^*) \leq L(\boldsymbol{z}, \boldsymbol{\alpha}_{\boldsymbol{z}})$.

You may follow the following steps but it is not required as long as your proof is correct.

1. Prove that $L(\boldsymbol{z}^*, \boldsymbol{\alpha}^*) \leq f(\boldsymbol{z}_p^*)$
2. Prove that $L(\boldsymbol{z}^*, \boldsymbol{\alpha}^*) \geq f(\boldsymbol{z}_p^*)$

Problem 4 (25 points) SVM Classification. Please refer to the Jupyter Notebook in the assignment, and complete the coding part in it! You can use sklearn SVM package: `https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC`