1. Consider the following training data set:

$$\mathbf{x}_1 = [1,0]^T, z_1 = -1$$
$$\mathbf{x}_2 = [0,1]^T, z_2 = -1$$
$$\mathbf{x}_3 = [0,-1], z_3 = -1$$
$$\mathbf{x}_4 = [-1,0]^T, z_4 = 1$$
$$\mathbf{x}_5 = [0,2]^T, z_5 = 1$$
$$\mathbf{x}_6 = [0,-2]^T, z_6 = 1$$
$$\mathbf{x}_7 = [-2,0]^T, z_7 = 1$$

Use following nonlinear transformation of the input vector $\mathbf{x} = [x_1, x_2]^T$ to the transformed vector $\mathbf{u} = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]^T$: $\varphi_1(\mathbf{x}) = x_2^2 - 2x_1 + 3$ and $\varphi_2(\mathbf{x}) = x_1^2 - 2x_2 - 3$. What is the equation of the optimal separating "hyperplane" in the $\mathbf{u}$ space?

2. Consider the following training data set :

$$\mathbf{x}_1 = [0,0]^T, z_1 = -1$$
$$\mathbf{x}_2 = [1,0]^T, z_2 = 1$$
$$\mathbf{x}_3 = [0,-1], z_3 = 1$$
$$\mathbf{x}_4 = [-1,0]^T, z_4 = 1$$

Note that in the following, you need to use equations that describe $\mathbf{w}$ and give rise to the dual optimization problem.

(a) Write down the dual optimization problem for training a Support Vector Machine with this data set using the polynomial kernel function

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^2$$

(b) Solve the optimization problem and find the optimal λ_i's using results about quadratic forms and check the results with Wolfram Alpha or any software package.

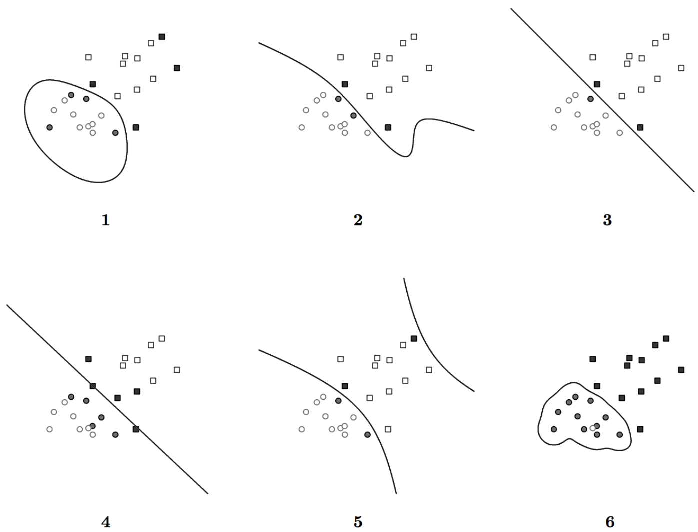
(c) Show that the equation of the decision boundary in a kernel SVM $\mathbf{w}^T\mathbf{u} + w_0 = 0$ can be represented as $g(\mathbf{x}) = \sum_{i=1}^N \lambda_i z_i \kappa(\mathbf{x}_i, \mathbf{x}) + w_0$.

(d) We learned that for vectors that do not violate the margin[1] (i.e. $z_j(\mathbf{w}^T\mathbf{u}_j + w_0) - 1 > 0$), the Lagrange multiplier is zero, i.e. $\lambda_j = 0$. On the other hand, for vectors on the margin ($z_j(\mathbf{w}^T\mathbf{u}_j + w_0) - 1 = 0$), $\lambda_j \neq 0$. Show that, consequently, one can find a vector $\mathbf{x}_j$ for which $\lambda_j \neq 0$ and calculate w_0 as $w_0 = 1/z_j - \sum_{i=1}^N \lambda_i z_i \kappa(\mathbf{x}_i, \mathbf{x}_j)$.

[1] For simplicity, consider Kernel SVM with hard margins, i.e. no slack variables.

(e) Sketch the decision boundary for this data set based on parts (2c) and (2d).

3. In the following figure, there are different SVMs with different decision boundaries. The training data is labeled as $z_i \in \{-1, 1\}$, represented as circles and squares respectively. Support vectors are drawn in solid circles. Determine which of the scenarios described below matches one of the 6 plots (note that one of the plots does not match any scenario). Each scenario should be matched to a unique plot. Explain your reason for matching each figure to each scenario.



(a) A soft-margin linear SVM with $C = 0.02$

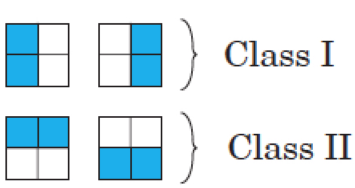
(b) A soft-margin linear SVM with $C = 20$

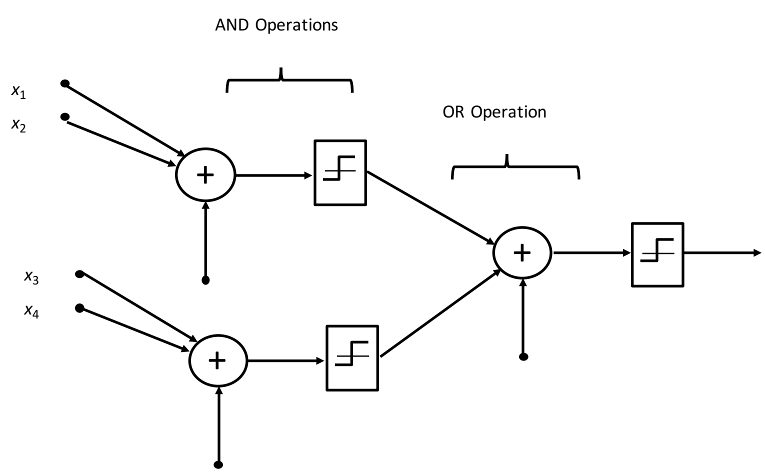
(c) A hard-margin kernel SVM with $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j + (\mathbf{x}_i^T\mathbf{x}_j)^2$

(d) A hard-margin kernel SVM with $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-5\|\mathbf{x}_i - \mathbf{x}_j\|^2)$

(e) A hard-margin kernel SVM with $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{1}{5}\|\mathbf{x}_i - \mathbf{x}_j\|^2)$

4. Consider the two classes of patterns that are shown in the figure below. Design a multilayer neural network with the following architecture to distinguish these categories.




AND Operations
OR Operation
x_1
x_2
x_3
x_4