CSCN8000 – Artificial Intelligence Algorithms and Mathematics

Assignment 1: Support Vector Machines

Question 1: [3 Points]

One of the most commonly used kernels in SVM is the Gaussian RBF kernel: $\mathrm{k}(x_i, x_j) = e^{\left(-\frac{|x_i - x_j|^2}{\{2\sigma^2\}}\right)}$

Suppose we have three points, z_1, z_2, and x. z_1 is geometrically very close to x, and z_2 is geometrically far away from x. What is the value of $\mathrm{k}(z_1, x)$and $\mathrm{k}(z_2, x)$?. **Choose one** of the following and **provide** an explanation of your choice by relating to what was said in class:

a) $\mathrm{k}(z_1, x)$ will be close to 1 and $\mathrm{k}(z_2, x)$ will be close to 0.
b) $\mathrm{k}(z_1, x)$ will be close to 0 and $\mathrm{k}(z_2, x)$will be close to 1.
c) $\mathrm{k}(z_1, x)$ will be close to c_1, $c_1 \gg 1$ and $\mathrm{k}(z_2, x)$will be close to c_2, $c_2 \ll 0$, where c_1, $c_2 \in$ R
d) $\mathrm{k}(z_1, x)$ will be close to c_1, $c_1 \ll 0$ and $\mathrm{k}(z_2, x)$will be close to c_2, $c_2 \gg 1$, where c_1, $c_2 \in$ R

Points:

z_1 : geometrically close to x

z_2 : far away from point x

For:

a) $\mathrm{k}(z_1, x)$ will be close to 1 and $\mathrm{k}(z_2, x)$ will be close to 0.

- $\mathrm{k}(z_1, x)$:
  - Since z_1 is very close to x, $||z_1 - x||^2$ is a small value.
  - Thus, the term $-\frac{|z_1 - x|^2}{\{2\sigma^2\}}$ will be a small negative number, making $k(z_1, x)$ close to $e^{(0)} = 1$
- $k(z_2, x)$:
  - Since z_2 is far from x, $||z_2 - x||^2$ is a large value.
  - The term $-\frac{|z_2 - x|^2}{\{2\sigma^2\}}$ will be large negative number, resulting in $k(z_2, x)$ approaching 0 (as $e^{(large\ negative\ value)}$ tends to vanish).
- The given option is valid since it accurately describes the behavior of the RBF kernel under the given conditions.

b) $k(z_1, x)$ will be close to 0 and $k(z_2, x)$ will be close to 1.

- $k(z_1, x)$:
  - This option incorrectly states that z_1 (which is close to x) would yield a kernel value close to 0. This is not correct because a small distance results in a kernel value close to 1.
- $k(z_2, x)$:

    - It incorrectly asserts that z_2 would give a kernel value close to 1 due to it being far from x. This is also not correct since large distance results in a kernel value approaching 0.
- This option is invalid due to the incorrect relationship between distance and kernel values.

c) $k(z_1, x)$ will be close to c_1, $c_1 \gg 1$, and $k(z_2, x)$ will be close to c_2, $c_2 \ll 0$ where c_1, $c_2 \in$ R

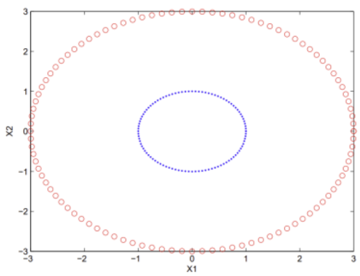
- $k(z_1, x)$:
    - While $k(z_1, x)$ is correctly assessed to be close to a high value, the statement that $c_1 \gg 1$ is misleading. The value should be close to 1, not necessarily greater than 1 significantly.
- $k(z_2, x)$:
    - The assertion that c_2 would be a value less than 0 is incorrect because kernel values for the RBF kernel are always non-negative (since $e^{(something\ negative)}$ is always positive).
- This option is invalid due to the incorrect characterizations of $k(z_1, x)$ being greater than 1 and $k(z_2, x)$ being less than 0.

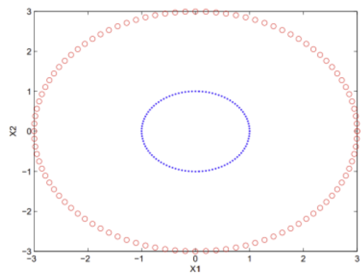
d) $k(z_1, x)$ will be close to c_1, $c_1 \ll 0$, and $k(z_2, x)$ will be close to c_2, $c_2 \gg 1$, where c_1, $c_2 \in$ R

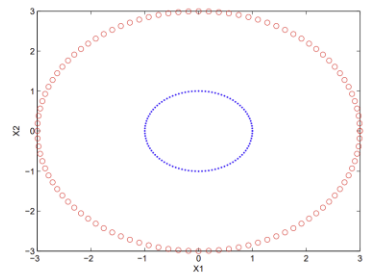
- $k(z_1, x)$:
    - This option incorrectly implies that $k(z_1, x)$ would approach a value less than 0. As previously discussed, the kernel outputs are always positive for the inputs, meaning $k(z_1, x)$ should be close to 1, not a negative value.
- $k(z_2, x)$:
    - This asserts that $k(z_2, x)$ would be a large value greater than 1, which again is incorrect. Kernel values for distances with substantial separation approach 0, not an excessively large value.
- This option is invalid due to the incorrect definitions of both kernel values in relation to their distances.

Question 2: [5 Points]

You are given the following 3 plots, which illustrates a dataset with two classes. **Draw the decision boundary** when you train an SVM classifier with linear, polynomial (order 2) and RBF kernels respectively. Classes have equal number of instances**. Provide a reasoning/explanation for your answer.**







1. Linear Kernel

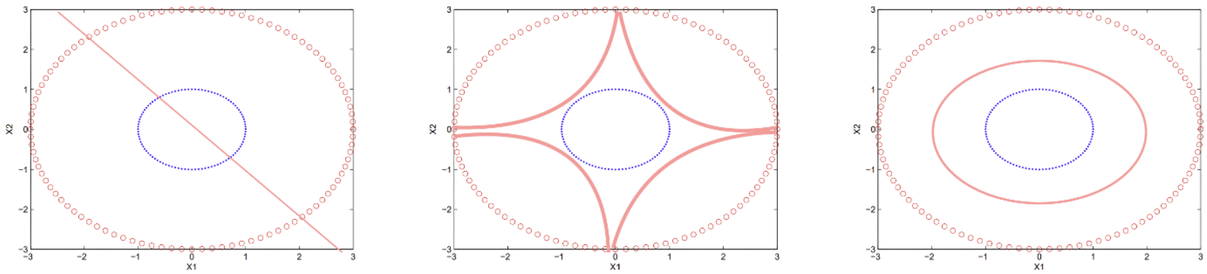
- A decision boundary can be a straight line or a hyperplane
- The linear kernel can only create linear boundaries, which won't effectively separate the inner and outer classes in patterns like concentric circles. This means the classification would be poor.

2. Polynomial Kernel (Order 2)

- Decision boundary of a polynomial kernel of order 2 can create a more complex decision boundary than a simple line, often resembling a curved line. However, unlike some other datasets, it may not perfectly match the circular shape of the classes.
- While it can capture some non-linear separability, in cases like this specific dataset, the polynomial kernel may not adapt as well as necessary to create a clear separation and might result in some points being misclassified. It could produce a boundary in the shape of a closed curve, but not necessarily a perfect fit to the circular pattern.

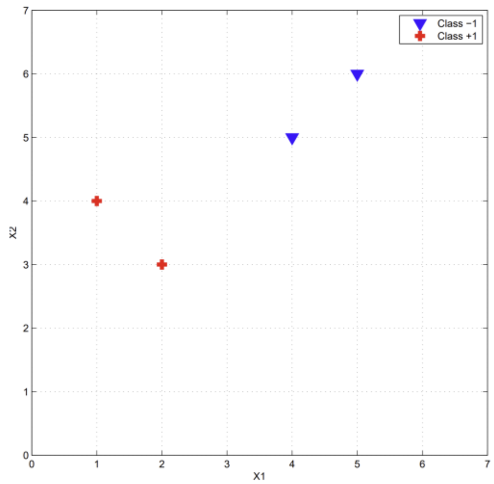
3. RBF Kernel

- We can se a highly flexible, smooth boundary that can adapt closely to the shapes of the data.
- The RBF kernel excels in situations where the data is not linearly separable by creating a decision boundary that can encircle the inner class, effectively separating it from the outer class. This capability allows it to perfectly fit the concentric circle pattern.



The linear kernel struggles with non-linear data, the polynomial kernel can provide a better fit but may not fully capture the circular formations, while the RBF kernel offers the necessary flexibility to adapt to the complex structure of the dataset, leading to a high level of accuracy in classification.

Question 3: [11 Points]

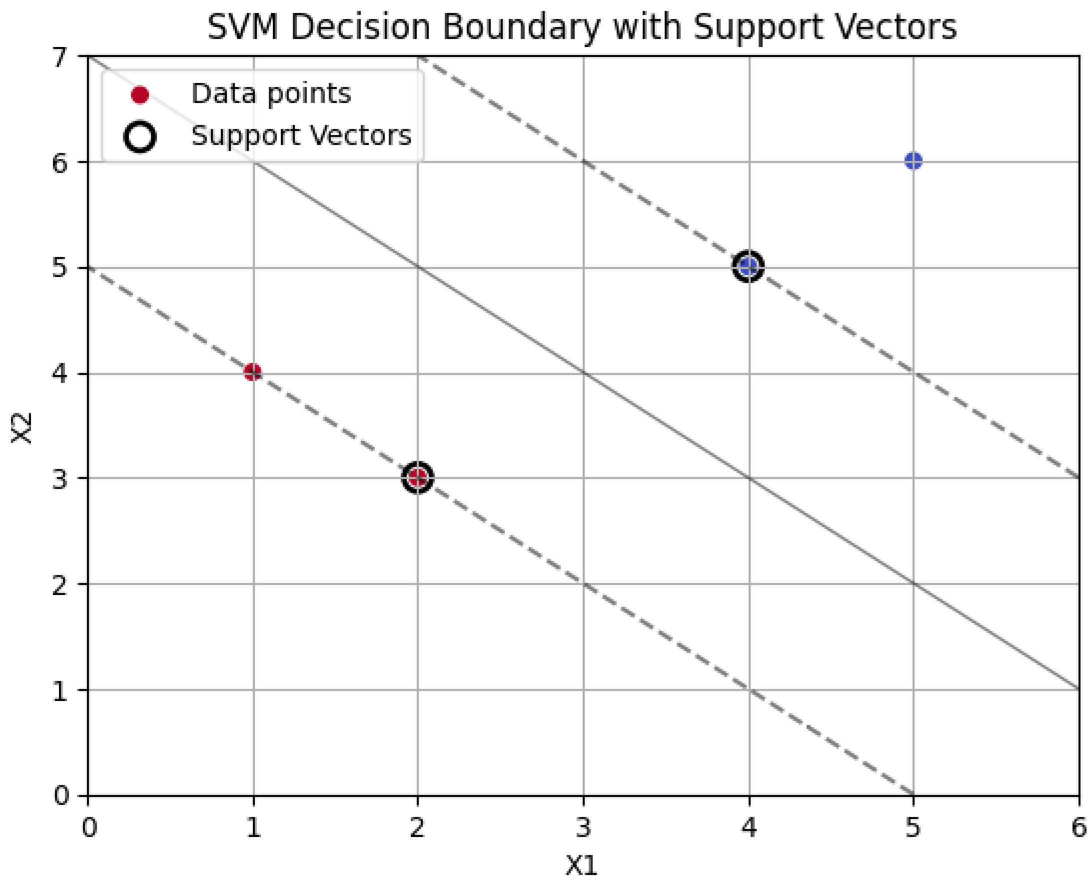


You are training a linear SVM on a tiny dataset with 4 points shown in Figure 2. This dataset consists of two examples with class label -1 (denoted with plus), and two examples with class label +1 (denoted with triangles).

a) **Find the weight vector $\boldsymbol{w}$ and bias $\boldsymbol{b}$**. What's the equation corresponding to the decision boundary? [7 Points]

   a. *Please note that a maximum of 2 grades will be given if only the final answer was provided without the detailed steps.*

b) **Circle** the support vectors and **draw** the decision boundary and the weight vector w. [4 Points]


SVM Decision Boundary with Support Vectors
Data points
Support Vectors
X1
X2
0
1
2
3
4
5
6
7

The equation corresponds to decision boundary: $x_1w_1 + x_2w_2 + b = 0$

Given points:

For class "red" denoted with class label 1 (denoted with plus) –

$$(1,4)$$

$$(2,3)$$

For class "blue" denoted with class label -1 (denoted with triangle) –

$$(4,5)$$

$$(5,6)$$

So, the equations for given points are:

For $(1,4)$:- $1w_1 + 4w_2 + b = 1$ --------------------------- eq 1

For $(2,3)$:- $2w_1 + 3w_2 + b = 1$ --------------------------- eq 2

For $(4,5)$:- $4w_1 + 5w_2 + b = -1$ --------------------------- eq 3

For $(5,6)$:- $5w_1 + 6w_2 + b = -1$ --------------------------- eq 4

STEP 1: eq 2 – eq 1

$2w_1 + 3w_2 + b - (1w_1 + 4w_2 + b) = 1 - 1$

$2w_1 + 3w_2 + b - 1w_1 - 4w_2 - b = 0$

$w_1 - w_2 = 0$

$w_1 = w_2$ --------------------------- eq 5

STEP 2: Putting in eq 6 in eq 2

$2w_1 + 3w_2 + b = 1$

$2w_1 + 3w_1 + b = 1$ From eq 5

$5w_1 + b = 1$ --------------------------- eq 6

STEP 3: Putting value of eq 5 in eq 3

$4w_1 + 5w_2 + b = -1$

$9w_1 + b = -1$ --------------------------- eq 7

STEP 4: eq 7 – eq 6

$9w_1 + b - (5w_1 + b) = -1 - (1)$

$9w_1 + b - \ 5w_1 - b = -2$

$4w_1 = -2$

$w_1 = -(\frac{1}{2})$ --------------------------- eq 8

Therefore, from eq 8

$w_2 = -\left(\frac{1}{2}\right)$ --------------------------- eq 9

STEP 5: put the value from eq 8 & 9 to eq 6

$5w_1 + b = \ 1$

$-\left(\frac{5}{2}\right) + b = \ 1$

$b = \ 1 + \left(\frac{5}{2}\right)$

$b = \ \left(\frac{7}{2}\right)$ --------------------------- eq 10

STEP 6: put the value from eq 8,9,10 to decision boundary equation which was:

$x_1w_1 + x_2w_2 + b = 0$

$-\left(\frac{1}{2}\right)x_1 - \left(\frac{1}{2}\right)x_2 + \left(\frac{7}{2}\right) = 0$

$\frac{-x_1 - x_2 + 7}{2} = 0$

$-x_1 - x_2 + 7 = 0$

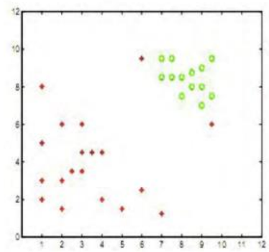
$x_1 + x_2 = 7$ --------------------------- eq 11

Therefore, $x_1 + x_2 = 7$ is an equation of a line as decision boundary

Also, as learnt from class that the support vectors contain the closest points to the decision boundary, separates the class. Moreover, looking at the datapoints the classes can be clearly separated with an even distribution.

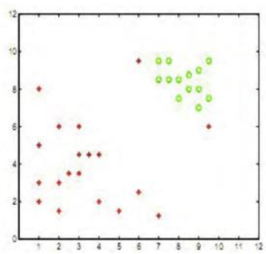
Question 4 [6 Points]:

Assume that we are training an SVM with a quadratic kernel - i.e. our kernel function is a polynomial kernel of degree 2. This means the resulting decision boundary in the original feature space may be parabolic in nature. The dataset on which we are training is given below:

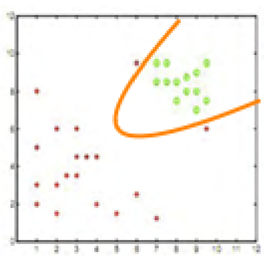


The slack penalty C will determine the location of the separating parabola. Please answer the following questions qualitatively.

a) [3 Points] Where would the decision boundary be for very large values of C? (Remember that we are using a quadratic kernel). **Justify your answer in one sentence and then draw the decision boundary in the figure below.**

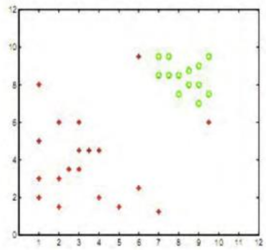
a. 

For very large values of *C*, the decision boundary will be positioned very close to the training samples of the green class while attempting to avoid misclassifying any red samples, effectively creating a complex parabolic curve that tightens around the green class.

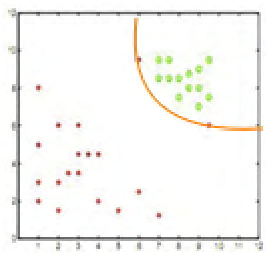


b) [3 Points] Where would the decision boundary be for C nearly equal to 0? **Justify your answer in one sentence and then draw the decision boundary in the figure below.**

a.

b. 

For CC nearly equal to 0, the decision boundary will be a broad parabolic curve that is less sensitive to the training data, resulting in potential misclassifications where the boundary does not tightly fit around the green points, allowing some red points to be included inside the region classified as green.



Deliverables

1. This is a written not a coding assignment. You're free to provide you answers to this assignment as handwritten scanned papers or electronically written documents (i.e. Word Document/Latex). Please make sure your submissions are neat, organized, and easy to read or it will be hard to fairly grade your assignment.
2. Submit a final scanned pdf named as "[Full Name]_[Student ID]_[Section Number]_Ass1.pdf"