Homework 3: SVM

(Due Date: Jul 1, 2025)

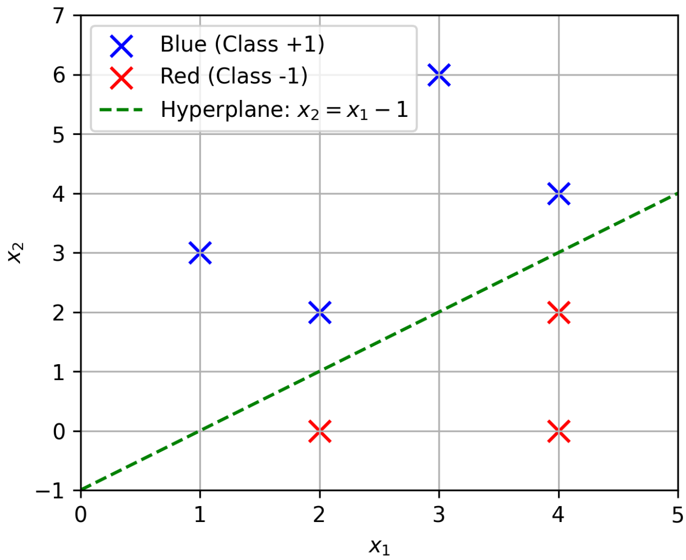
MATH QUESTIONS

Question 1

1.

We are given a set of 7 observations in $\mathbb{R}^2$, each labeled as either **Blue** or **Red**. The 7 points are plotted, as shown in Figure 1. It is observed that the two classes are linearly separable. The Red class (label -1) and the Blue class (label $+1$) can be separated by a straight line. Among all possible separating lines, the optimal separating hyperplane is the one that maximizes the margin between the two classes and passes equidistantly between the nearest opposing class points, known as support vectors. The equation of the hyperplane is $x_2 = x_1 - 1$ and is explained in details in part 2.

Figure 1: SVM separating hyperplane $x_2 = x_1 - 1$ that maximally separates the two classes.



2.

The classification rule for a support vector classifier is defined using the decision function:

$$f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Training-Time Constraint:

For the optimal separating hyperplane in a linearly separable case, the training data must satisfy the following margin constraints:

$$y_i f(\mathbf{x}_i) = y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \geq 1$$

This implies:

$$f(\mathbf{x}_i) = \begin{cases} +1 \text{ (Blue class)} & \text{if } f(\mathbf{x}_i) \geq 1 \\ -1 \text{ (Red class)} & \text{if } f(\mathbf{x}_i) \leq -1 \end{cases}$$

Only the points that lie exactly at these margins are the *support vectors*.

Hyperplane Coefficients:

Visually, we could see from Figure 1 that the midpoint between (2,2) (blue class) and (2,0) (red class) is (2,1). Also, the midpoint between (4,4) (blue class) and (4,2) (red class) is (4,3). The linear hyperplane (shown in green dashed line in Figure 1) must pass through the midpoints (2,1) and (4,3). In addition, the support vectors are:

- Blue class: $(2, 2)$ and $(4, 4)$
- Red class: $(2, 0)$ and $(4, 2)$

The optimal separating hyperplane lies equidistant between these support vectors and is defined by the equation:

$$x_2 = x_1 - 1 \quad \text{or equivalently,} \quad f(x) = -x_1 + x_2 + 1 = 0$$

Here, the parameters of the hyperplane are:

$$\beta_0 = 1, \quad \beta_1 = -1, \quad \beta_2 = 1$$

Testing-Time Classification Rule:

At test time, we classify a new input $\mathbf{x}$ using the sign of the decision function:

$$\hat{y} = \text{sign}(f(\mathbf{x})) = \begin{cases} +1 & \text{if } f(\mathbf{x}) > 0 \\ -1 & \text{if } f(\mathbf{x}) < 0 \end{cases}$$

This rule assigns the point to the Blue class if it lies above the hyperplane, and to the Red class if it lies below.

3.

The margin ρ of the maximal margin hyperplane is defined as the perpendicular distance between the closest training points (support vectors) and the separating hyperplane.

$$\rho = \frac{2}{\|\boldsymbol{\beta}\|} \quad \text{where} \quad \|\boldsymbol{\beta}\| = \sqrt{\beta_1^2 + \beta_2^2}$$

From Part 2, we have:

$$\beta_1 = -1, \quad \beta_2 = 1$$

Thus, the norm of the weight vector is:

$$\|\boldsymbol{\beta}\| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}$$

Substituting into the margin formula:

$$\rho = \frac{2}{\sqrt{2}} = \sqrt{2} \approx 1.414$$

Therefore, the margin of the maximal margin hyperplane is:

$$\rho = \sqrt{2}$$

This value represents the total distance between the closest point of each class and the decision boundary. Each support vector lies exactly $\frac{\sqrt{2}}{2}$ units away from the hyperplane.

4. Support vectors are the training points that lie exactly on the margin boundaries of the classifier. These satisfy the condition:

$$y_i f(\mathbf{x}_i) = 1 \quad \text{or equivalently} \quad f(\mathbf{x}_i) = \begin{cases} +1 & \text{if } y_i = +1 \\ -1 & \text{if } y_i = -1 \end{cases}$$

where the decision function is:

$$f(x) = -x_1 + x_2 + 1$$

Evaluating this function on the dataset, the following points satisfy the support vector condition:

- Blue class (label $+1$):
  - $(2, 2)$: $f(x) = -2 + 2 + 1 = 1$
  - $(4, 4)$: $f(x) = -4 + 4 + 1 = 1$

- Red class (label -1):
  - $(2,0)$: $f(x) = -2+0+1 = -1$
  - $(4,2)$: $f(x) = -4+2+1 = -1$

Thus, the support vectors are:

$$(2,2),\ (4,4),\ (2,0),\ (4,2)$$

5.

Observation 7 is the point $(4,0)$ from the Red class (label -1).
The decision function is:

$$f(x) = -x_1 + x_2 + 1$$

Evaluating for point 7:

$$f(4,0) = -4+0+1 = -3$$

This value is far from the margin boundary for the Red class, which lies at $f(x) = -1$. Therefore, point 7 lies well outside the margin region and is not a support vector.

According to the theory of SVM, only the support vectors (the training points closest to the decision boundary) determine the position of the maximal-margin hyperplane. Points that are not support vectors do not influence the margin as long as they remain outside the margin boundary.

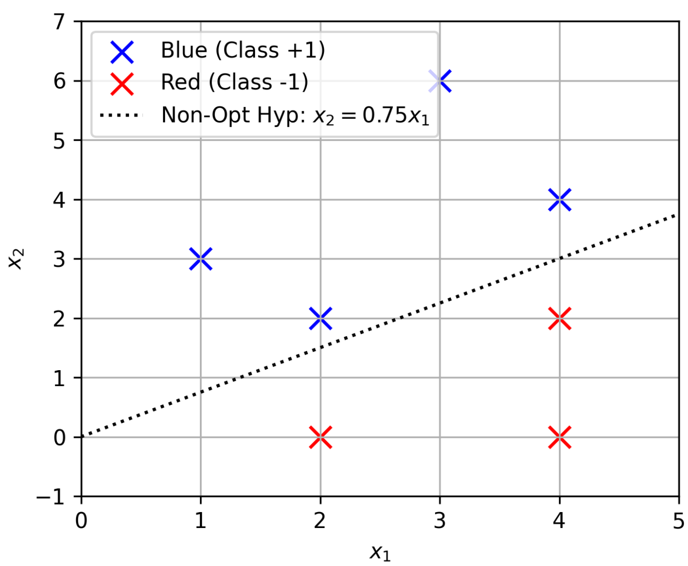
Therefore, a slight movement of the seventh observation $(4,0)$ will not affect the maximal margin hyperplane because it is not a support vector and remains outside the margin.

6.

An alternative separating hyperplane, depicted in Figure 2, is one that correctly separates the two classes but does not achieve the maximum margin. For example, consider the line:

$$x_2 = 0.75x_1 \quad \text{or equivalently,} \quad f(x) = -0.75x_1 + x_2 = 0$$

Figure 2: An alternate separating non-optimal hyperplane $x_2 = 0.75x_1$.



We evaluate this function for all training points:

- **Blue class (label $+1$):**
  - $(3,6):\ f(x) = \frac{27}{4}$
  - $(2,2):\ f(x) = 0.5$

- $(4,4): \ f(x) = 4$
- $(1,3): \ f(x) = \frac{9}{4}$

- **Red class (label –1):**
  - $(2,0): \ f(x) = -1.5$
  - $(4,2): \ f(x) = -1$
  - $(4,0): \ f(x) = -3$

All points are classified correctly based on the sign of $f(x)$. However, this hyperplane does not lie equidistant between the closest points from each class, and hence, the margin is smaller than that of the optimal hyperplane.

Therefore, the hyperplane defined by $f(x) = -0.75x_1 + x_2 = 0$ separates the two classes, but it is not the optimal separating hyperplane.

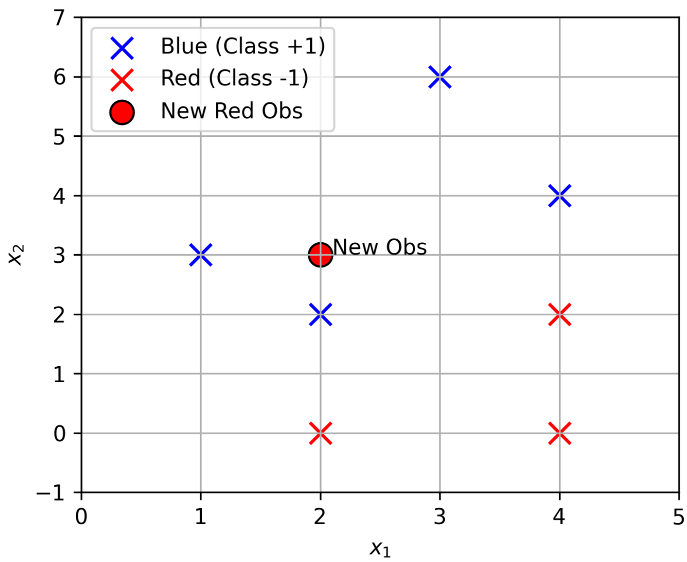
7.

To make the two classes no longer linearly separable, we can add a new observation that introduces class overlap across the existing margin.

Let us add the following point:

$$(2,3) \quad \text{with label } -1 \text{ (Red class)}$$

Figure 3: Adding an additional observation destroys linear separability of the two classes.



This new point lies within the region currently occupied by the Blue class (label +1), as shown in Figure 3

By introducing a Red-labeled point into this Blue region, it becomes impossible to draw a single straight line that correctly separates all Red and Blue points. As a result, the classes become non-linearly separable.

Question 2

1.

We are given a dataset with 4 samples, each consisting of 2 features and a binary class label. The points are as follows:

- Positive class $(+1)$: $(1, 1)$ and $(-1, -1)$
- Negative class (-1): $(1, -1)$ and $(-1, 1)$

The data set can be represented in the following table:

Table 1: Training Dataset.

Index	x_1	x_2	y
1	1	1	$+1$
2	-1	-1	$+1$
3	1	-1	-1
4	-1	1	-1

The feature matrix X has the the shape 4×2 (4 samples, 2 features), and the label vector y has shape 4×1:

$$X \in \mathbb{R}^{4 \times 2}, \quad y \in \mathbb{R}^4$$

Bonus (Logic Gate Representation): The class label y is $+1$ when $x_1 = x_2$, and -1 when $x_1 \neq x_2$. Table 2 below represents the representation of the XNOR gate with SVM input (± 1) (or the inverse of the exclusive OR operation).

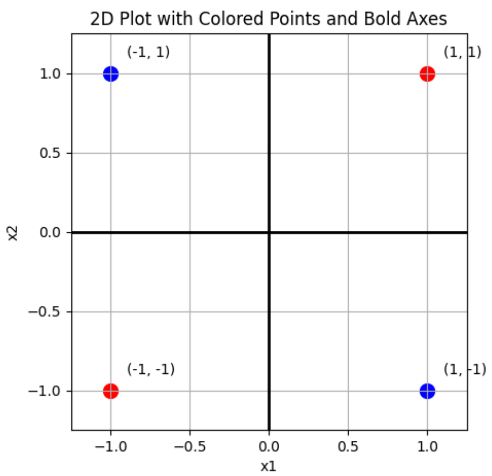
Table 2: XNOR Gate Representation with SVM Inputs (± 1)

x_1	x_2	Output (y)
$+1$	$+1$	$+1$
-1	-1	$+1$
$+1$	-1	-1
-1	$+1$	-1

2.

The four data points are plotted, shown in Figure 4, in the two-dimensional x_1–x_2 plane using their coordinates and labels. The plot reveals that the two classes form a cross-like pattern. Points with the same label lie diagonally opposite to one another, making it impossible to draw a single straight line that separates the positive class from the negative class.

Figure 4: Plot of the points with red and blue colors representing data from classes 1 and -1, respectively.



3.

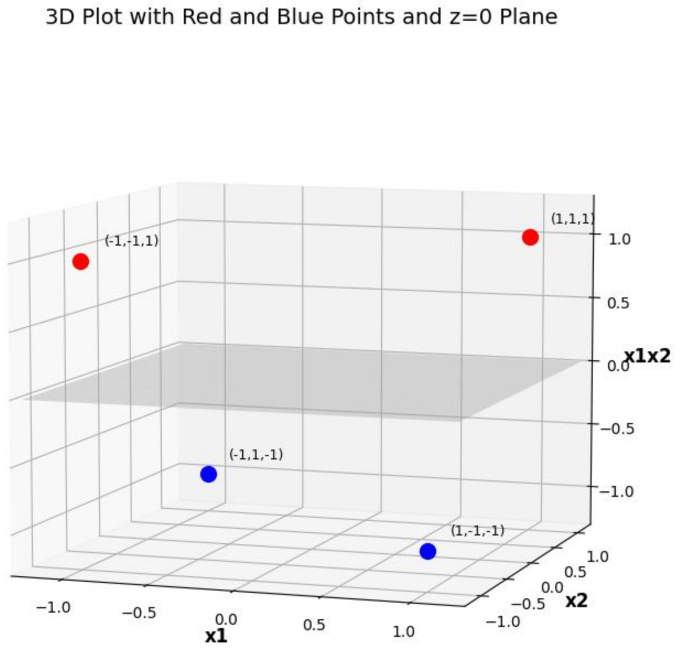
Let us apply the given non-linear feature transformation:

$$\phi(x) = [x_1,\ x_2,\ x_1 x_2]$$

The transformed data set becomes:

Point	x_1	x_2	x_1x_2	Label (y)
(1, 1)	1	1	1	+1
(−1, −1)	−1	−1	1	+1
(1, −1)	1	−1	−1	−1
(−1, 1)	−1	1	−1	−1

Figure 5: Plot of the transformed data points in 3-Dimensional space.



As clearly visible from the 3D plot (shown in Figure 5), the transformation $x_1,\ x_2,\ x_1x_2$ renders the positive and negative classes in clearly linearly separable manner.

4.

After the transformation $\phi(x) = [x_1, x_2, x_1x_2]$, the two classes become linearly separable in $\mathbb{R}^3$ using only the third coordinate.

We define the classifier using:

$$f(\phi(x)) = x_1x_2 \quad \Rightarrow \quad \beta_1 = 0,\ \beta_2 = 0,\ \beta_3 = 1,\ \beta_0 = 0$$

The margin is calculated as:

$$\rho = \frac{2}{\|\boldsymbol{\beta}\|} = \frac{2}{\sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2}} = \frac{2}{\sqrt{1}} = 2$$

All four data points lie at equal (unit) distance from the separating hyperplane $x_1x_2 = 0$.

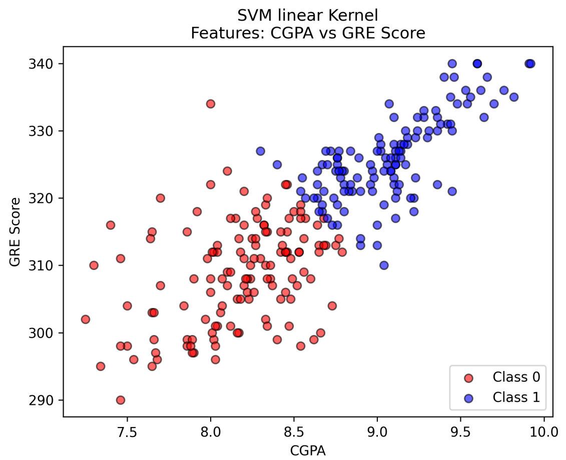
- $(1, 1)$ and $(-1, -1)$ from the positive class
- $(1, -1)$ and $(-1, 1)$ from the negative class

PROGRAMMING QUESTIONS

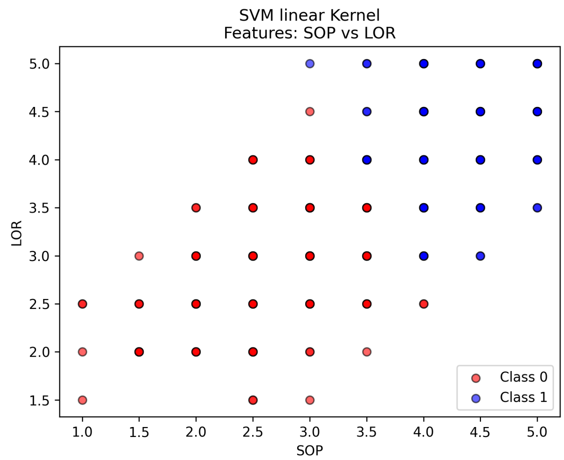
Question 3

(e) Result Visualization

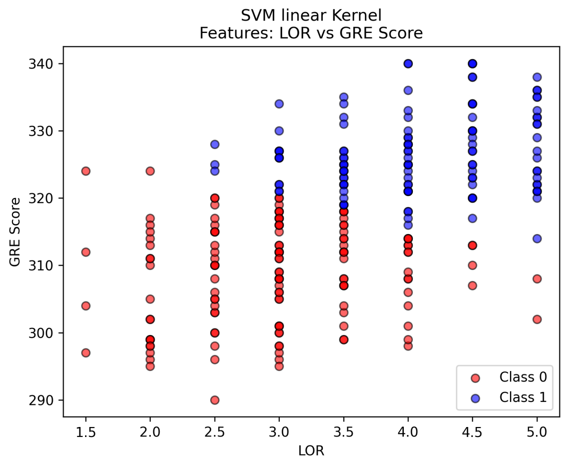
Figure 6: SVM Prediction - Linear Kernel Visualizations for Different Feature Pairs



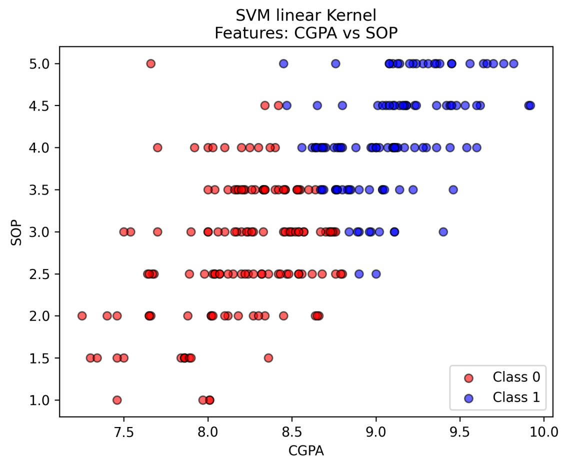
(a) CGPA vs GRE Score



(b) SOP vs LOR

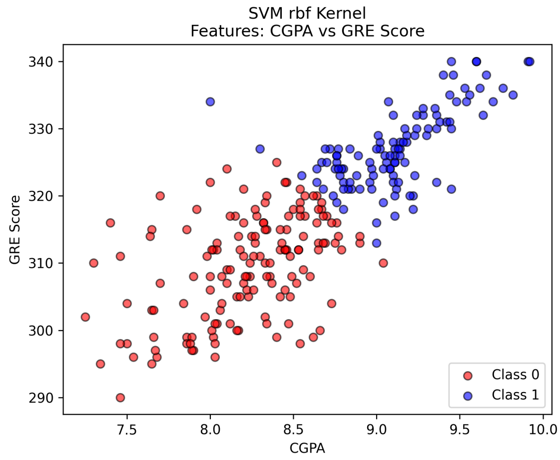


(c) LOR vs GRE Score

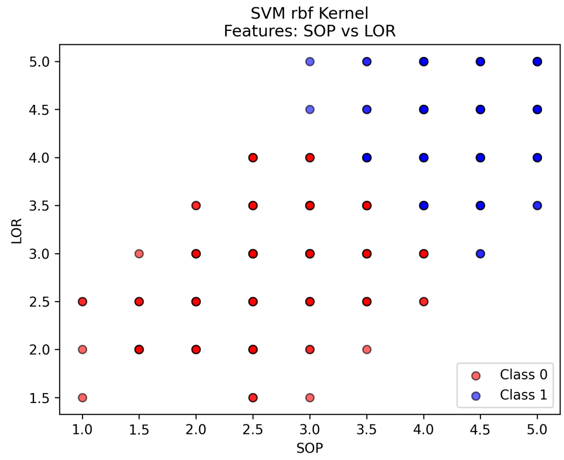


(d) CGPA vs SOP

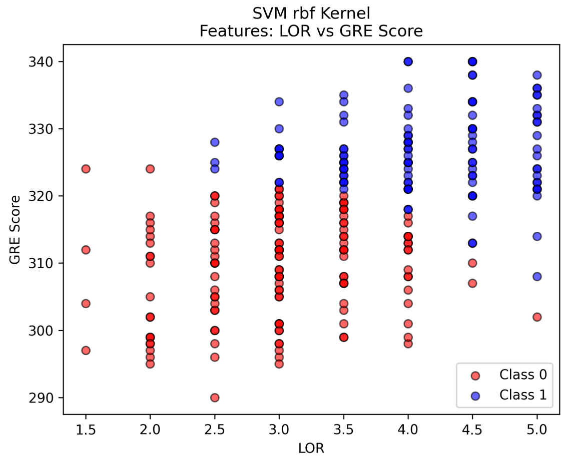
Figure 7: SVM Prediction - RBF Kernel Visualizations for Different Feature Pairs



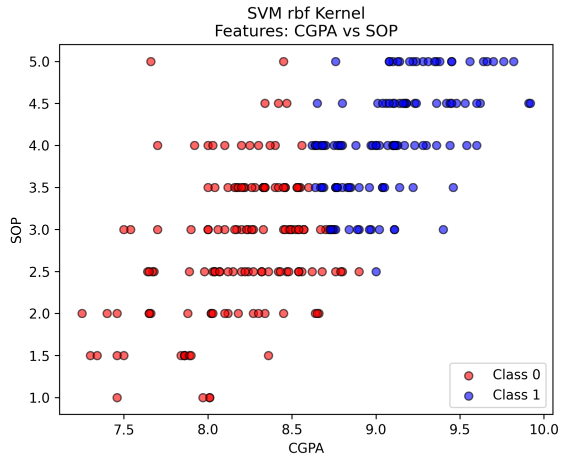
(a) CGPA vs GRE Score



(b) SOP vs LOR

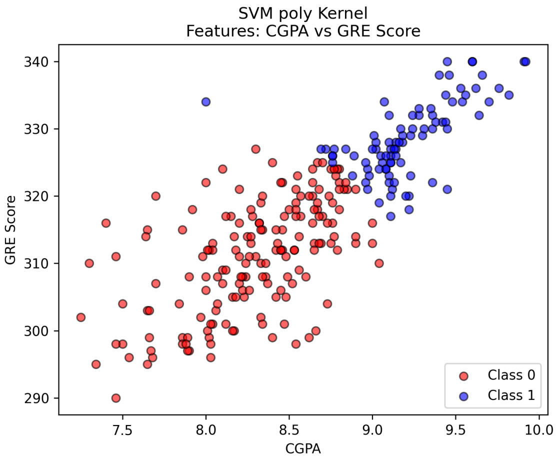


(c) LOR vs GRE Score

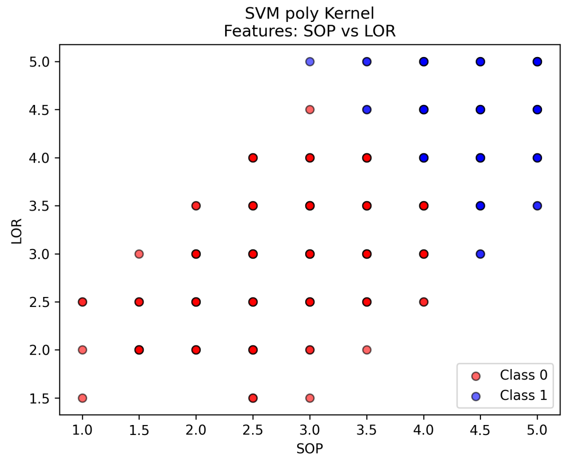


(d) CGPA vs SOP

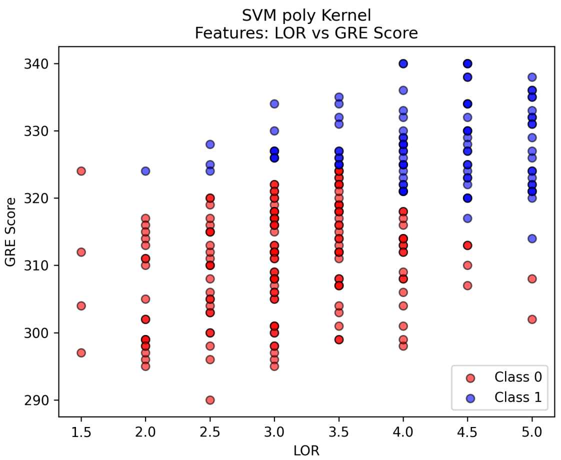
Figure 8: SVM Prediction - Polynomial (degree 3) Kernel Visualizations for Different Feature Pairs



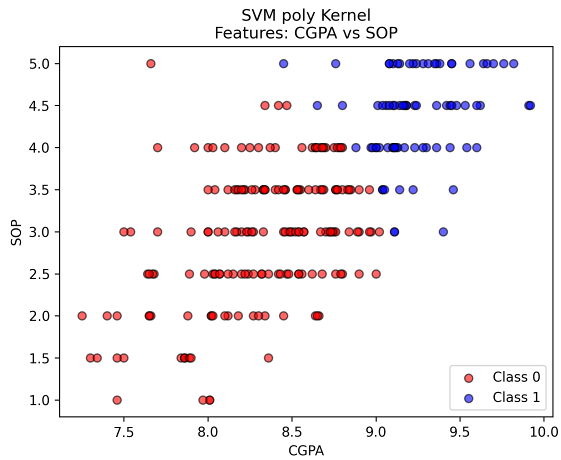
(a) CGPA vs GRE Score



(b) SOP vs LOR



(c) LOR vs GRE Score



(d) CGPA vs SOP

(e) Result Analysis

Based on the visualizations, the best performance on the training set is achieved with the RBF kernel using CGPA and GRE Score as features. This combination shows the clearest separation between classes with minimal overlap and smooth decision boundaries.