Instructions for homework submission

1. Complete two parts in this homework:
   - Math questions: Include your solution in LaTeX document. Show your work. Submission with embedded photos of handwritten work will not be graded.
   - Programming questions: Complete the given skeleton `Python` code. **For questions requiring visualization, analysis, and discussion, please include your solution in the same LaTeX document.**
2. Submit your work to Gradescope including:
   - A PDF document for written parts: `FirstName_LastName_HW3.pdf`. LaTeX source code is not required.
   - A completed `Python` code: `FirstName_LastName_HW3.py`.
   - **There are two separate submission portals on `Gradescope`:** one for code and one for the report. Submitting your work to the wrong portal will result in a loss of marks.
   - Please assign your answer in PDF report to its corresponding question when submitting to Gradescope. Submitting your work without assigning correspoinding question will result in a loss of marks.
3. Start early!
4. Total: 100 points.

Math Questions

Question 1 (30 points)

We are given $n = 7$ observations in $p = 2$ dimensions. Each observation has an associated class label.

Index	X_1	X_2	Y
1	3	6	Blue
2	2	2	Blue
3	4	4	Blue
4	1	3	Blue
5	2	0	Red
6	4	2	Red
7	4	0	Red

1. Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.
2. Describe the classification rule for the maximal margin classifier. Provide values for β_0, β_1, and β_2.
3. Indicate the margin for the maximal margin hyperplane.
4. Identify the support vectors for the maximal margin classifier.
5. Explain if a slight movement of the seventh observation affects the maximal margin hyperplane.
6. Draw an alternative hyperplane that is not the optimal separating hyperplane, and provide its equation.
7. Add an additional observation so that the two classes are no longer separable by a hyperplane.

Question 2 (20 points)

We have a training dataset with 4 samples, 2 features, and 2 classes. The positive examples are $(1, 1)$ and $(-1, -1)$. The negative examples are $(1, -1)$ and $(-1, 1)$.

1. Draw a table representing this training set. What is the shape of X and y? (Bonus: Identify the logic gate represented by this table.)
2. Plot the points on the x-y plane. Are these points linearly separable?
3. Apply the feature transformation $\phi(x) = [x_1, x_2, x_1x_2]$ and plot the transformed points. Are these transformed points linearly separable?
4. Determine the margin size after the transformation and identify the support vectors.

Programming Questions

Question 3 (50 points)

We will use university application data to classify admissions.

(a) Data Pre-processing

1. Create a binary label named `label` based on the column `Chance of Admit`. Assign 1 if the value is greater than the median, otherwise 0.
2. Apply appropriate pre-processing to data.
3. Split the data into training and validation sets (80-20 split). Submission with data leakage will result in a loss of marks.

(b) Model Initialization

Initialize 3 different SVM models with the following kernels:

1. Linear kernel
2. RBF kernel
3. Polynomial (degree 3) kernel

(c) Feature Selection and Model Training

Train each SVM model with these feature combinations:

- CGPA and SOP
- CGPA and GRE Score
- SOP and LOR
- LOR and GRE Score

(d) Support Vectors

Identify the support vectors for each model and feature combination.

(e) Result Visualization

Visualize the predictions for each kernel-feature combination on the training set. Use color coding for labels. Include this figure in your report.

(f) Result Analysis

Determine the best feature-kernel combination based on training set figures. Validate the model on the validation split and find the best performing combination with respect to accuracy. Initialize `my_best_model` variable using your best model combination. Tune hyperparameter and aim for 0.83 accuracy on our reserved test set.