**Instructions**: Please work in your preassigned groups to complete and submit your work to the appropriate folder in LumiNUS. Please submit two documents:

1. Completed word file named as Group-X-Activity-Week-4.docx and
2. Your working ipynb file named as Group-X-Activity-Week-4.ipynb
3. Consider building an SVM for the following two-class training data:

Positive class: (-1, 3) (0, 2) (0, 1) (0, 0), Negative class: (1, 5) (1, 6) (3, 3)

1. Plot the training points and, by inspection, draw a linear classifier that separates the data with maximum margin.
2. The linear SVM is parameterized by h(x) = (w^t)(x) + b. What are the parameters w and b for this problem?
3. Suppose you observe an additional set of points, all from the positive class. What is the linear SVM (in terms of **w** and b) now?
4. More positive points: (−2, 0) (−2, 1) (−2, 3) (−1, 0) (−1, 1) (0, 0)
5. Consider the metrics accuracy, precision, and recall.
6. Give one reason why each of the metrics listed above is useful.
7. Suppose you are given the same test dataset and two binary classifiers. Give a simple example such that Classifier 1 has higher accuracy than Classifier 2, but Classifier 2 has both higher precision and higher recall than Classifier 1? Hint: Give 2x2 confusion matrix for each classifier.
8. In this problem, we will look at the Breast Cancer Wisconsin (Diagnostic) Data Set available UCI Machine Learning Repository. Please use the wdbc.data dataset from:

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29

Compute the performance of the SVM algorithm on this dataset for predicting the whether the cancer is malignant or benign. Use a random train/test data split of 70%/30%. Repeat this process 20 times and compute the average performance.

Please evaluate the following algorithms:

* SVM1: SVM with linear kernel
* SVM2: SVM with RBF kernel
* SVM3: Same as SVM2 but with regularization (soft margin), vary C and report your best results.

Please compute the following metrics and fil in the table below.

* Training Accuracy and Test Accuracy
* Precision and Recall (which are important metrics that complement Accuracy)
* You can read about performance metrics at: <https://en.wikipedia.org/wiki/Confusion_matrix>
* SKLearn contains functions to compute these metrics:

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | | Precision | Recall |
|  | Train | Test |  |  |
| SVM1 |  |  |  |  |
| SVM2 |  |  |  |  |
| SVM3  C = |  |  |  |  |