School of Computer Science Engineering and Technology
Assignment-07

Course- B.Tech	**Type-** Core
Code-23CS106	**Course Name-** Artificial Intelligence & Machine Learning
Year- 2024-2025	**Semester-** Even, **Instructor:** Prof. E.L.N. Kiran
Date– 30-01-2024	**Batch-** AIML-A,B

Implement Support Vector Machine Classification using Breast Cancer Dataset

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. This gap is also called maximum margin and the SVM classifier is called maximum margin clasifier.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

1. Import the Libraries required for SVM. [CO2]

Import all libraries required along with visualization to completed the task on SVM

2. Import the Breast Cancer Dataset from Sklearn Packages. [CO1]

- Once downloaded the Breast Cancer data, prepare the dictionary format to access the data using its keys.
- Describe all the features from the dictionary, feature names.
- Setup the dataframe , describe itd details, check for missing values.
- Identify the target class and assign it to the dataframe.
- Perform exploratory analysis of dataframe using `seabon - sns package`
- the Draw the boxplot of first 10 columns to verify their role in cancer.

3. Train and Test Data . [CO3]

- Prepare the Train and Test data from the dataframe.
- Drop the `cancer` columns and define the dataframe with only `target` results
- Split the data into train, test using `train_test_split`

4. Train the SVC using the Train Dataset. [CO3]

- `import sklearn.svm import SVC`
- Apply the `model.fit` to dataset

5. Predict and Analysis the Performance of the SVC Model. [CO4]

- Apply `model.predict(X_test)`
- Generate the classification_report, confusion_matrix using `sklearn.metrics`

6. Improve the Accuracy of Model using GridSearchCV Model. [CO4]

- Given the following parameters validate the GridSearchCV Model:
  `param_grid = 'C': [0.1,1, 10, 100, 1000], 'gamma': [1,0.1,0.01,0.001,0.0001], 'kernel': ['rbf']`
- Implement the model.predict
- Generate the classification_report, confusion_matrix using `sklearn.metrics`