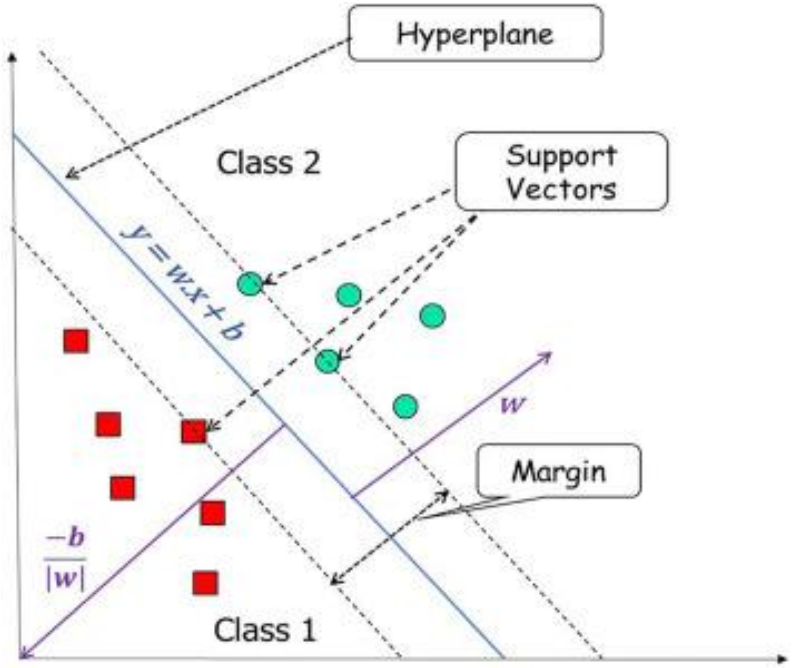
CT60A4800 Fundamentals of smart systems – Assignment 4

Name: Trieu Huynh Ba Nguyen Student number: 000405980

The Iris flower dataset was used for this simple machine learning model. It is a is a multivariate data set created by British biologist and statistician Ronald Fisher. The data set contains 50 samples of 3 species of Iris, and each of them has the length and the width of the sepals and petals.

According to a research topic by Brownlee (2019), the Support Vector Machines (SVM) model proved to be the most accurate. Therefore, in this assignment, this model will be replicated for the Iris flower dataset.

In machine learning, SVMs are supervised learning models, and can analyze data for both classification and regression analysis. The goal of the SVM formula is to make the most effective line or call boundary which will segregate n-dimensional area into categories in order that we will simply place the new datum within the correct class in the future. This best decision boundary is named a hyperplane. SVM chooses the acute points/vectors that facilitate in making the hyperplane. These extreme cases are called as support vectors, and thus algorithm is termed as Support Vector Machine. contemplate the below diagram during which there are 2 totally different classes that are classified employing a decision boundary or hyperplane:



Source: Poonia, Singh & Nayak (2022)

This solution to the assignment has many limitations, due to the inadequacy in my own knowledge and skills when it comes to advanced statistics and techniques for machine learning. I do not have the ability to interpret the results either.

Python code:

```
# Load necessary libraries
from pandas import read_csv
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.svm import SVC

# Load Iris dataset
dataset = read_csv(
    "https://raw.githubusercontent.com/jbrownlee/Datasets/master/iris.csv
", names=['sepal-length', 'sepal-width', 'petal-length', 'petal-width',
'class'])

# Split-out validation dataset
array = dataset.values
X = array[:, 0:4]
y = array[:, 4]
X_train, X_validation, Y_train, Y_validation = train_test_split(
    X, y, test_size=0.20, random_state=1)

# Make predictions on validation dataset
model = SVC(gamma='auto')
model.fit(X_train, Y_train)
predictions = model.predict(X_validation)

# Evaluate predictions
print(accuracy_score(Y_validation, predictions))
print(confusion_matrix(Y_validation, predictions))
print(classification_report(Y_validation, predictions))
```

Result:

```
0.9666666666666667
[[11  0  0]
 [ 0 12  1]
 [ 0  0  6]]
                 precision    recall  f1-score   support

    Iris-setosa       1.00      1.00      1.00        11
Iris-versicolor       1.00      0.92      0.96        13
 Iris-virginica       0.86      1.00      0.92         6

       accuracy                           0.97        30
      macro avg       0.95      0.97      0.96        30
   weighted avg       0.97      0.97      0.97        30
```

Video: Link to OneDrive

References

Brownlee, J. (2019). *Your First Machine Learning Project in Python Step-By-Step*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/machine-learning-in-python-step-by-step/.

Poonia R. C., Singh, V. and Nayak, S. R. (2022). *Deep Learning for Sustainable Agriculture*. Academic Press.