Assignment 2 : Support Vector Machine

Part 1 : Soft SVM using quadratic programming

In this soft SVM using quadratic programming we use the cvxopt solvers for solving the convex problem which is to maximise ½*x^t*p*x+q^t*x such that g*x<=h and ax=b.
In the above formulation we replace p by graham matrix which is y*y*x*x , q=-1 , g = - 1, h=0 , b=0 , a=y^t , x= alphas which is to be maximised.

In this I have used the data set which is provided in file.csv which is of **Room Occupancy** can be found at link:https://www.kaggle.com/sachinsharma1123/room-occupancy.
To run the Program, first upload the data set provided with the .ipynb file or download from the above mentioned site.
Description of Dataset:

1. Experimental data used for **binary classification** (room occupancy) from Temperature,Humidity,Light and CO2.
2. This dataset contains 5 features and a target variable: Temperature, Humidity, Light, Carbon dioxide(CO2)
3. Target Variable: Occupancy
    - 1-if there is a chance of room occupancy.
    - -1-No chances of room occupancy (initially 0, set to -1 during processing)
4. This Dataset contains 2666 rows and 6 columns with 5 features and 1 target.
5. No missing values.

I have used the CVXOPT Solvers. All the required steps are already written in the .ipynb file, Different Splits are also Provided; Maximum accuracy is found as 98%.

Part 2 : Soft SVM using Stochastic Gradient Descent

In this i have used the algorithm:
For i in iterations:

```
if(1-y(<w,x>-b)<=0):
        w-=-2/c*w
        b=0
else:
        w-= -2/c*w-yi*xi
        b-=-yi
```

Hence we update the weight and bias using the learning rate and the regularization parameters.
In this I have used the data set which is provided in file.csv which is of **Room Occupancy** can be found at link:https://www.kaggle.com/sachinsharma1123/room-occupancy.
To run the Program, first upload the data set provided with the .ipynb file or download from the above mentioned site.
Description of Dataset:

1. Experimental data used for **binary classification** (room occupancy) from Temperature,Humidity,Light and CO2.
2. This dataset contains 5 features and a target variable: Temperature, Humidity, Light, Carbon dioxide(CO2)
3. Target Variable: Occupancy
   - 1-if there is a chance of room occupancy.
   - -1-No chances of room occupancy (initially 0, set to -1 during processing)
4. This Dataset contains 2666 rows and 6 columns with 5 features and 1 target.
5. No missing values.

I have coded the Soft SVM and optimised using the Stochastic gradient descent rule. All the required steps are already written in the .ipynb file, Different Splits are also Provided; Maximum accuracy is found as 97% on test data and 97% on train data.

Part 3 : Hard SVM using quadratic programming

In this soft SVM using quadratic programming we use the cvxopt solvers for solving the convex problem which is to maximise ½*x^t*p*x+q^t*x such that g*x<=h and ax=b.
In the above formulation we replace p by graham matrix which is y*y*x*x , q=-1 , g = - 1, h=0 , b=0 , a=y^t , x= alphas which is to be maximised also w=yi*alphaI*xi.

In this I have used the data set which is provided in IRIS.csv which is of **IRIS Flower dataset**, and can be found at link:https://www.kaggle.com/arshid/iris-flower-dataset.
To run the Program, first upload the data set provided with the .ipynb file or download from the above mentioned site.
Description of Dataset:

1. Experimental data used for **binary classification** (flower species classification) from sepal_length,sepal_width,petal_length and petal_width.
2. This dataset contains 4 features and a target variable: Temperature, Humidity, Light, Carbon dioxide(CO2)
3. Target Variable: Species
   - 1-for one kind of species.
   - -1-for other kind
4. This Dataset contains 100 rows and 5 columns with 4 features and 1 target.
5. No missing values.

I have used the CVXOPT Solvers. All the required steps are already written in the .ipynb file, Different Splits are also Provided; Maximum accuracy is found as 95%.

Note:

1. I have also used randomly created Linearly separable data to show the working of svm graphically.