

## 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  $\frac{1}{2}x^t p x + q^t x$  such that  $g^t x \leq h$  and  $Ax = b$ .

In the above formulation we replace  $p$  by gram matrix which is  $y^t y x x^t$ ,  $q = -1$ ,  $g = -1$ ,  $h = 0$ ,  $b = 0$ ,  $A = y^t$ ,  $x = \text{alphas}$  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  $(1 - y(\langle w, x \rangle - b) \leq 0)$ :

$w = -2/c * w$

$b = 0$

else:

$w = -2/c * w - y_i * x_i$

$b = -y_i$

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  $\frac{1}{2}x^t p x + q^t x$  such that  $g^t x \leq h$  and  $ax = b$ .

In the above formulation we replace  $p$  by gram matrix which is  $y^t y x x^t$ ,  $q = -1$ ,  $g = -1$ ,  $h = 0$ ,  $b = 0$ ,  $a = y^t$ ,  $x = \alpha$  which is to be maximised also  $w = y_i \alpha_i x_i$ .

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:

6. Experimental data used for **binary classification** (room occupancy) from Temperature, Humidity, Light and CO2.
7. This dataset contains 5 features and a target variable: Temperature, Humidity, Light, Carbon dioxide(CO2)
8. 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)
9. This Dataset contains 2666 rows and 6 columns with 5 features and 1 target.
10. 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 64%.