CS360: Lab Assignment 6

SVM

There are two datasets:

- *Data.csv*
- Cancer dataset (folder contains *Training data.csv* and *Testing data.csv*)

The last column of each dataset represents class labels (+1 or -1) corresponding to the row, ie., each row represents features of one data point plus class label. Use the attached matlab files to do the programs

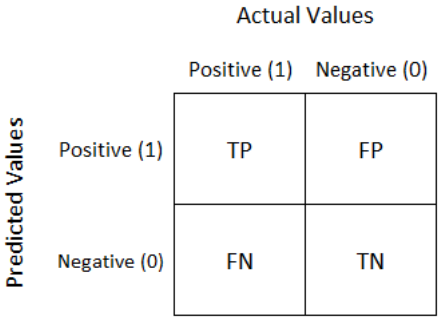
In $Q.1$, you can start with *svm1.m* in which *Data.csv* is used for training. Here we used Sequential Minimal Optimization (SMO) tool to get α_i's which will be used to compute weight vector $\mathbf{w}$ and bias b as follows:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$
$$b = \frac{1}{m} \sum_i (y_i - \mathbf{w}^T \mathbf{x}_i)$$

where $\mathbf{x}_i$ and y_i denote i^{th} input vector and corresponding label, m is the total number of input samples. You can use the same dataset for the prediction in which,

$$\hat{y_i} = \begin{cases} +1, & \text{if } (\mathbf{w}^T \mathbf{x}_i + b) > 0 \\ -1, & \text{otherwise} \end{cases}$$

After getting $\hat{y_i}$, we need to compute confusion matrix from which accuracy and F1 score can be calculated. You need to write code in *confusionMatrix.m* which will return confusion matrix, accuracy and F1 score.



$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

$$\text{F1} = \left(\frac{\text{Recall}^{-1}+\text{Precision}^{-1}}{2}\right)^{-1}$$

In *Q*.2, you can move to *svm2.m* in which the cancer dataset is used for the experiment. The folder *cancer dataset* contains *Training data.csv* and *Testing data.csv* which can be used for training and testing. As described in the Q.1, you can compute accuracy and F1 score.