Assignment 2 : Support Vector Machine **Rajat Jain**
University of Central Florida UCF ID : 4446187
Prof : Dr. Gita Sukthankar Due Date: Oct 8,2018
TA: Navid Kardan

Problem Statement

Apply SVM on glass dataset four kernels Linear,Polynomial,Sigmoid and RBF. Using One vs One, One vs Rest and One vs One with class weight approach. Perform comparison between all the approaches implemented.

What is SVM ?

Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.SVM constructs hyperplane or set of hyperplanes in a high or infinite-dimensional space which can be used for classification, regression,etc. Good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class,the larger the margin the lower the generalization error of the classifier.

Hyperparameters and its effect on SVM

SVM has set of parameters which are called as hyperparameters.

1. Soft margin constant C : is a regularization parameter that controls the trade off between the achieving a low training error and a low testing error that is the ability to generalize your classifier to unseen data. If C is too large the optimization algorithm will try to reduce weight as much as possible leading to a hyperplane which tries to classify each training example correctly. Doing this will lead to loss in generalization properties of the classifier. On the other hand if C is too small then you give your objective function a certain freedom to increase weight a lot, which will lead to large training error.

2. Kernel Parameters :

   - Gamma : is parameter of gaussian kernel to handle non-linear classification. It controls the tradeoff between error due to bias and variance in the model.
   - Degree of the polynomial kernel : The lowest degree polynomial is the linear kernel, which is not sufficient when a non-linear relationship between feature exists. Higher degree kernel can map patterns into higher dimensional space and therefore yield a higher dimensional hyperplane.
   - Coef0 (r) : Its the parameter for kernel projection.

So the parameters for kernels are:

1. Linear : C

2. RBF : C, Gamma

3. Sigmoid: C, Gamma, Coef0

4. Polynomial: C, Gamma, Coef0, Degree

Range tried in this assignment are as follows:

- C = [math.pow(2,-5),math.pow(2,-3),math.pow(2,-1), math.pow(2,0),math.pow(2,1),math.pow(2,3)]
- Gamma = [math.pow(2,-15), math.pow(2,-13),math.pow(2,-11),math.pow(2,-9),math.pow(2,-7),math.pow(2,-5),math.pow(2,-3)]
- Coef0 also called as r = [0,0.1,0.01,0.001]
- Degree =[2,3,4]
- But for sigmoid I increased range of c and decreased range of gamma as these range were not giving good accuracy
- Range of C for Sigmoid is : [math.pow(2,-5),math.pow(2,-3),math.pow(2,-1), math.pow(2,0),math.pow(2,1 math.pow(2,8), math.pow(2,11), math.pow(2,13), math.pow(2,15)]
- Range of gamma for Sigmoid is : [math.pow(2,-15), math.pow(2,-13),math.pow(2,-11)]
- I was trying these range in rbf, polynomial , and linear but was not getting good accuracy, so i removed them from testing params so that execution time to find optimal hyperparam decreases and you can execute my code in less time

Implementation Steps

Steps for Assignment Point 1:

1. Read the data from glass.data file
2. Created function for linear svm, rbf svm, poly svm and sigmoid svm
3. For One vs One approach, passed parameter decision function shape="ovo"
4. Apply cross validation, kept 1 fold for test and divided rest in 80 perecent for training and 20 percent for validation
5. Created multiple model using different sets of hyperparameter and got the accuracy using validation set.
6. Found the optimal hyperparameter i.e. the one with highest accuracy. Used those hyperparameter on test set we left and got the accuracy for that fold
7. Repeated step 4,5,6 for all folds and later averaged all the accuracy which we got from each fold and got the final accuracy for One vs One Classifier

Steps for Assignment Point 2:

1. Changed the parameter decision function shape to "ovr" i.e. one vs rest. And repeated all the steps as done in Assignment Point 1

Steps for Assignment Point 3:

1. Comparision for one vs one and one vs rest is shown in next sections.

Steps for Assignment Point 4:

1. Used parameter class weight="balanced" in svm and repeated all the steps as done in part 1

Accuracy, Hyperparameters and Execution Time

One vs One Classifier

1. Linear :
   - Accuracy : 64.28 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.6428571428571429 Using optimal hyperparameter C as 8.0
     (b) Final Accuracy for fold 2 is 0.5952380952380952 Using optimal hyperparameter C as 8.0
     (c) Final Accuracy for fold 3 is 0.5952380952380952 Using optimal hyperparameter C as 1.0
     (d) Final Accuracy for fold 4 is 0.5714285714285714 Using optimal hyperparameter C as 8.0
     (e) Final Accuracy for fold 5 is 0.8095238095238095 Using optimal hyperparameter C as 1.0
   - Execution time =0.8 seconds. Note this is total time to find optimal parameter and using optimal parameter on on test set
2. Sigmoid:
   - Accuracy : 63.33 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.7142857142857143 Using Optimal hyperparameter C as 32768.0 gamma as 3.0517578125e-05 and r as 0
     (b) Final Accuracy for fold 2 is 0.6428571428571429 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0.1
     (c) Final Accuracy for fold 3 is 0.5952380952380952 Using Optimal hyperparameter C as 32768.0 gamma as 3.0517578125e-05 and r as 0
     (d) Final Accuracy for fold 4 is 0.6190476190476191 Using Optimal hyperparameter C as 8192.0 gamma as 0.0001220703125 and r as 0
     (e) Final Accuracy for fold 5 is 0.5952380952380952 Using Optimal hyperparameter C as 8192.0 gamma as 0.0001220703125 and r as 0
   - Execution time =1.105 seconds. Note this is total time to find optimal parameter and using optimal parameter on on test set
3. Polynomial:
   - Accuracy : 67.14 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.5476190476190477 Using Optimal hyperparameter C as 0.5 gamma as 0.125 r as 0.001 and degree as 3
     (b) Final Accuracy for fold 2 is 0.6904761904761905 Using Optimal hyperparameter C as 0.03125 gamma as 0.125 r as 0 and degree as 3

    - (c) Final Accuracy for fold 3 is 0.7380952380952381 Using Optimal hyperparameter C as 0.03125 gamma as 0.0078125 r as 0.1 and degree as 4
    - (d) Final Accuracy for fold 4 is 0.7142857142857143 Using Optimal hyperparameter C as 0.5 gamma as 0.03125 r as 0.01 and degree as 4
    - (e) Final Accuracy for fold 5 is 0.6666666666666666 Using Optimal hyperparameter C as 0.03125 gamma as 0.03125 r as 0.001 and degree as 4
  - Execution time=2 minutes and 61 seconds. Note this is total time to find optimal parameter and using optimal parameter on on test set

4. Rbf:
   - Accuracy : 69.52 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     - (a) Final Accuracy for fold 1 is 0.6428571428571429 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     - (b) Final Accuracy for fold 2 is 0.7142857142857143 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     - (c) Final Accuracy for fold 3 is 0.6190476190476191 Using Optimal hyperparameter C as 8.0 gamma as 0.03125
     - (d) Final Accuracy for fold 4 is 0.7142857142857143 Using Optimal hyperparameter C as 1.0 gamma as 0.125
     - (e) Final Accuracy for fold 5 is 0.7857142857142857 Using Optimal hyperparameter C as 8.0 gamma as 0.125
   - Execution time =0.3 seconds. Note this is total time to find optimal parameter and using optimal parameter on on test set
5. ***Conclusion***, from above data we can say that rbf works best for us as its accuracy is best and execution time is less. Also polynomial is giving good accuracy but its little less than rbf and execution time is huge

One vs Rest Classifier

1. Linear :
   - Accuracy : 64.30 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     - (a) Final Accuracy for fold 1 is 0.6190476190476191 Using optimal hyperparameter C as 1.0
     - (b) Final Accuracy for fold 2 is 0.6438571428571429 Using optimal hyperparameter C as 8.0
     - (c) Final Accuracy for fold 3 is 0.5952380952380952 Using optimal hyperparameter C as 1.0
     - (d) Final Accuracy for fold 4 is 0.5952380952380952 Using optimal hyperparameter C as 2.0
     - (e) Final Accuracy for fold 5 is 0.6904761904761905 Using optimal hyperparameter C as 8.0
   - Execution time =0.7 seconds. Note this is total time to find optimal parameter and using optimal parameter on on test set

2. Sigmoid:

   - Accuracy : 64.76 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.7142857142857143 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0
     (b) Final Accuracy for fold 2 is 0.7142857142857143 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0
     (c) Final Accuracy for fold 3 is 0.5238095238095238 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0
     (d) Final Accuracy for fold 4 is 0.6190476190476191 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0
     (e) Final Accuracy for fold 5 is 0.6666666666666666 Using Optimal hyperparameter C as 32768.0 gamma as 3.0517578125e-05 and r as 0
   - Execution time = 1.08 seconds. Note this is total time to find optimal parameter and using optimal parameter on on test set

3. Polynomial:

   - Accuracy : 70.4 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.6904761904761905 Using Optimal hyperparameter C as 0.03125 gamma as 0.125 r as 0.01 and degree as 3
     (b) Final Accuracy for fold 2 is 0.7142857142857143 Using Optimal hyperparameter C as 0.125 gamma as 0.03125 r as 0.01 and degree as 4
     (c) Final Accuracy for fold 3 is 0.7142857142857143 Using Optimal hyperparameter C as 0.03125 gamma as 0.03125 r as 0.01 and degree as 4
     (d) Final Accuracy for fold 4 is 0.6190476190476191 Using Optimal hyperparameter C as 0.125 gamma as 0.0078125 r as 0.1 and degree as 3
     (e) Final Accuracy for fold 5 is 0.7857142857142857 Using Optimal hyperparameter C as 0.5 gamma as 0.03125 r as 0.01 and degree as 3
   - Execution time= 2 minutes and 65 seconds

4. Rbf:

   - Accuracy : 70.47 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.7857142857142857 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     (b) Final Accuracy for fold 2 is 0.6190476190476191 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     (c) Final Accuracy for fold 3 is 0.8333333333333334 Using Optimal hyperparameter C as 2.0 gamma as 0.125
     (d) Final Accuracy for fold 4 is 0.6428571428571429 Using Optimal hyperparameter C as 8.0 gamma as 0.03125
     (e) Final Accuracy for fold 5 is 0.6428571428571429 Using Optimal hyperparameter C as 2.0 gamma as 0.125

- Execution time= 0.3 seconds. Note this is total time to find optimal parameter and using optimal parameter on test set

5. ***Conclusion***, from above data we can say that rbf works best for us as its accuracy is best and execution time is less. Also polynomial is giving almost same as rbf but execution time is huge

One vs One Classifier with class weight

1. Linear :
   - Accuracy : 62.85 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.4523809523809524 Using optimal hyperparameter C as 8.0
     (b) Final Accuracy for fold 2 is 0.6428571428571429 Using optimal hyperparameter C as 0.5
     (c) Final Accuracy for fold 3 is 0.6904761904761905 Using optimal hyperparameter C as 2.0
     (d) Final Accuracy for fold 4 is 0.5714285714285714 Using optimal hyperparameter C as 1.0
     (e) Final Accuracy for fold 5 is 0.6190476190476191 Using optimal hyperparameter C as 8.0
   - Execution time = 0.6 seconds.Note this is total time to find optimal parameter and using optimal parameter on test set
2. Sigmoid:
   - Accuracy :57.61 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.4523809523809524 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0
     (b) Final Accuracy for fold 2 is 0.5952380952380952 Using Optimal hyperparameter C as 8192.0 gamma as 0.0001220703125 and r as 0
     (c) Final Accuracy for fold 3 is 0.6190476190476191 Using Optimal hyperparameter C as 8192.0 gamma as 0.0001220703125 and r as 0
     (d) Final Accuracy for fold 4 is 0.6428571428571429 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0
     (e) Final Accuracy for fold 5 is 0.5714285714285714 Using Optimal hyperparameter C as 32768.0 gamma as 0.0001220703125 and r as 0.001
   - Execution time =1.11 seconds. Note this is total time to find optimal parameter and using optimal parameter on test set
3. Polynomial:
   - Accuracy : 62 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.6666666666666666 Using Optimal hyperparameter C as 0.03125 gamma as 0.03125 r as 0.1 and degree as 4

    (b) Final Accuracy for fold 2 is 0.5714285714285714 Using Optimal hyperparameter C as 0.125 gamma as 0.125 r as 0.01 and degree as 3
    (c) Final Accuracy for fold 3 is 0.6666666666666666 Using Optimal hyperparameter C as 0.03125 gamma as 0.03125 r as 0.1 and degree as 4
    (d) Final Accuracy for fold 4 is 0.6428571428571429 Using Optimal hyperparameter C as 1.0 gamma as 0.03125 r as 0.01 and degree as 4
    (e) Final Accuracy for fold 5 is 0.5476190476190477 Using Optimal hyperparameter C as 0.125 gamma as 0.03125 r as 0.01 and degree as 4
  - Execution time= 2minutes 25 seconds .Note this is total time to find optimal parameter and using optimal parameter on test set

4. Rbf:
   - Accuracy : 65.71 percent
   - Optimal Hyperparameter for each fold with accuracy is :
     (a) Final Accuracy for fold 1 is 0.6428571428571429 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     (b) Final Accuracy for fold 2 is 0.6428571428571429 Using Optimal hyperparameter C as 2.0 gamma as 0.125
     (c) Final Accuracy for fold 3 is 0.7142857142857143 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     (d) Final Accuracy for fold 4 is 0.5952380952380952 Using Optimal hyperparameter C as 8.0 gamma as 0.125
     (e) Final Accuracy for fold 5 is 0.6904761904761905 Using Optimal hyperparameter C as 8.0 gamma as 0.125
   - Execution time=0.45 seconds .Note this is total time to find optimal parameter and using optimal parameter on test set

5. ***Conclusion***, from above data we can say that rbf works best for us as its accuracy is best and execution time is less. Also polynomial and linear performed almost same but execution time for poly is too huge

Comparison between different classifier implemented.

1. ***Comparison One vs One and One vs Rest***: In both cases RBF works out the best and Sigmoid is worst among all. Acurracy of polynomial is almost same as RBF but it's execution time is huge. So if you want to choose between Polynomial and Linear than you have to think of tradeoff between time and accuracy because linear runs way faster than polynomial but accuracy is less than polynomial.
   Also if compare one vs one to one vs rest, one vs rest is faster and gives us better accuracy

2. ***Comparison One vs One and One vs One with Class Weight***: For given data set One vs One is performing better compare to One vs One with class weight. Execution time is almost same but accuracy we received is very low compare to one vs one. But with executing both classifier, RBF turns out to be the best in all