Advanced Kernel Methods - Assignment 3

Ragja Palakkadavath - SC18M003

January 2019

1. **Solution:**

   (a) (i) SVM via Iterative Method Results:

   (a) Data 1:
   RMSE obtained is 0.30171595227611786
   C obtained after cross validation is 1000
   Best epsilon obtained after cross validation is 1e-06
   Best kernel is Gaussian Kernel

   (b) Data 2:
   RMSE obtained is 0.07975847872284317
   C obtained after cross validation is 1000
   Best epsilon obtained after cross validation is 1e-06
   Best kernel is Linear

   (c) Data 3:
   RMSE obtained is 2.7977583701921644
   C obtained after cross validation is 1000
   Best epsilon obtained after cross validation is 1e-06
   Best kernel is Gaussian

   (ii) SVM via LibSVM(SMO Regression) in weka Results:

   (a) Data 1:
   RMSE obtained is 0.0007
   Best kernel is Polynomial Kernel

   (b) Data 2:
   RMSE obtained is 0.0372
   Best kernel is Gaussian

   (c) Data 3:
   RMSE obtained is 3.2406
   Best kernel is Polynomial

   (iii) SVM via Kernel Ridge Regression Results:

(a) Data 1:
RMSE obtained is 0.2979489399437001
Best kernel is Gaussian
Best lambda value is 1

(b) Data 2:
RMSE obtained is 0.08806516901933743
Best kernel is Gaussian
Best lambda value is 1

(c) Data 3:
RMSE obtained is 4.429909497199123
Best kernel is Polynomial
Best lambda value is 10000

(iv) SVM via Kernel Online Ridge Regression Results:

(a) Data 1:
RMSE obtained is 0.1768818585521945
Learning Rate obtained after cross validation is 1
Best kernel is Hyperbolic
Best lambda value is 0.0001

(b) Data 2:
RMS obtained is 0.07429717033732061
Learning Rate obtained after cross validation is 1
Best kernel is Hyperbolic
Best lambda value is 0.0001

(c) Data 3:
RMSE obtained is 4.861393975198768
Learning Rate obtained after cross validation is 1
Best kernel is Hyperbolic
Best lambda value is 0.0001

(b) T-Test was performed between Iterative Method and Kernel Online Ridge Regression. p value was found to be 0.4687. They are significantly(95%) giving the same results. T-Test was performed between Kernel Ridge Regression and Kernel Online Ridge Regression. p value was found to be 0.0485. They are significantly(95%) giving the same results.

(c) Batch Learning and Online Learning gives almost same performance based on the accuracy values.

2. **Solution:**

(a) (i) SVM via Classification Iterative Method Results:

(a) Data 4:
Accuracy obtained is 0.9
C obtained after cross validation is 100000

Best kernel is Gaussian

(b) Data 6: Accuracy obtained is 0.9
C obtained after cross validation is 1000
Best kernel is Hyperbolic

(ii) SVM via SMO Classification Results:

(a) Data 4: Accuracy obtained is 1.0
C obtained after cross validation is 0.01
Best kernel is Linear

(b) Data 6:
Accuracy obtained is 0.5 - Not converging for all values of C parameters
C obtained after cross validation is 0.0001
Best kernel is Hyperbolic

(iii) SVM via Chunking Method Results:

(a) Data 4:
Accuracy obtained is 0.8666666666666667
C obtained after cross validation is 0.0001
Best kernel is Hyperbolic

(b) Data 6:
Accuracy obtained is 0.9
C obtained after cross validation is 0.0001
Best kernel is Hyperbolic

(iv) SVM via Decomposition Results:

(a) Data 4:
Accuracy obtained is 0.8666666666666667
C obtained after cross validation is 0.01
Best kernel is Hyperbolic
Best working set size is 8

(b) Data 6:
Accuracy obtained is 0.9
C obtained after cross validation is 0.01
Best kernel is Hyperbolic
Best working set size is 12

(b) Stopping Criteria used is the KKT conditions being satisfied for all the data points in the set.

3. **Solution:**

(a) KPCA on Data 5 - SVM Classification

(b) New points are present in $data5\_kpca\_hyperbolic.csv$

(i) KPCA was applied on Data 5 using the kernels: Gaussian, Polynomial with Degree 2, Hyperbolic and Linear

(ii) Best Accuracy was obtained when the new points obtained after PCA was run for **Hyperbolic Kernel**
Duality GAP is: 0.00023013823327484373
Dual Objective Function value is: -5925.10576646687
Primal Objective Function value is: -5925.1059966051025
Dual Slack variable value is 1.1071167751009404e-06
Primal Slack variable value is 2.58043494727427e-10
Value of bias is -0.05053189731436236
Accuracy obtained is 0.9333333333333333
C obtained after cross validation is 10000

(c) KPCA on Data 3 - SVM Regression

(d) New points are present in *data3_kpca_poly.csv*

(i) KPCA was applied on Data 3 using the kernels: Gaussian, Polynomial with Degree 2, Hyperbolic and Linear

(ii) Best RMSE was obtained when the new points obtained after KPCA was run for **Polynomial Kernel**
RMSE: 2.0411

4. **Solution:**

(a) SVM Classification Algorithm applied on Data5
Results:
Duality GAP is: 1.7410741325246767e-09
Dual Objective Function value is: -4.93632090734645e-06
Primal Objective Function value is: -4.938061981478532e-06
Dual Slack variable value is 8.609285358236643e-14
Primal Slack variable value is 1.0775081806594421e-05
Value of bias is -178.76542855982916
Accuracy obtained is 0.9333333333333333
C obtained after cross validation is 0.0001
Best kernel is P

(b) SVM Classification Algorithm applied on support vectors obtained from Data 5
Duality GAP is: 8.402277596228242e-09
Dual Objective Function value is: -4.933840900456772e-06
Primal Objective Function value is: -4.941811632612854e-06
Dual Slack variable value is 1.0443703693095454e-12
Primal Slack variable value is 7.15101564361333e-05
Value of bias is -173.61345466951664
Accuracy obtained is 1.0
C obtained after cross validation is 0.0001
Best kernel is P

(c) The performance of SVM algorithm applied on the support vectors obtained from Data 5 gives an accuracy of 1.0 whereas the performance of SVM algorithm directly on the Data set no. 5 gives only 0.93. Therefore, running the algorithm on the support vectors obtained from the data gives better results from this example.

5. **Solution:**

$$\alpha_{10} = -\eta * (f_9(x_{10}) - y_{10})$$
$$\alpha_5 = \alpha_i * (1 - \eta\lambda)^{5-i}$$

6. **Solution:**
Adult Data : SVM SMO using WEKA software:

(a)

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.947    0.487    0.860      0.947    0.901      0.531    0.730     0.854     class0
                 0.513    0.053    0.754      0.513    0.610      0.531    0.730     0.504     class1
Weighted Avg.    0.842    0.383    0.834      0.842    0.831      0.531    0.730     0.770

=== Confusion Matrix ===

     a     b   <-- classified as
 23405  1315 |     a = class0
  3820  4021 |     b = class1
```

Figure 1: Accuracy Metrics for Adult Data

(b) C obtained after cross validation is 1

(c) Best kernel is Linear

(d) Learning rate obtained after cross validation is 0.001

7. **Solution:**
Iris Data : multi class SVM (3 classes: Iris Setosa, Iris VersiColor, Iris Virginica)

(a) Results for running multi class SVM on Iris Data(old points):
Average accuracy obtained after cross validation(hold out): 0.9526315789473685

(b) KPCA was run on Iris Data set for the kernels: Linear, Polynomial with degree 2, Gaussian and Hyperbolic Kernels.

(c) New data points have been stored in the output directory as csv files

(d) Multi Class SVM applied on modified points of KPCA:

   (i) Results obtained for Linear Kernel:
   Average accuracy obtained after cross validation(hold out): 0.9421052631578949

   (ii) Results obtained for Polynomial Kernel:
   Average accuracy obtained after cross validation(hold out): 0.9157894736842105

   (iii) Results obtained for Gaussian Kernel:
   Average accuracy obtained after cross validation(hold out): 0.7052631578947369

   (iv) Results obtained for Hyperbolic Kernel:
   Average accuracy obtained after cross validation(hold out): 0.9473684210526315

8. **Solution:**
Let A be a symmetric matrix.

Let u and v be eigenvectors of A corresponding to any two distinct eigenvalues of A, say μ and λ.

Therefore we have Au=μu and Av=λv.

To prove that u and v are orthogonal, we have to show that $\langle u, v\rangle = 0$.

Keeping this in mind, $\mu\langle u, v\rangle = \langle \mu u, v\rangle = \langle Au, v\rangle = (Au)^T v = u^T A^T v = u^T Av = u^T \lambda v = \lambda\langle u, v\rangle$

Hence we obtain, $\mu\langle u, v\rangle = \lambda\langle u, v\rangle$.

$(\mu - \lambda)\langle u, v\rangle = 0$

Since $\mu$and $\lambda$are distinct eigenvalues of A, $(\mu - \lambda) \neq 0$. Therefore, $\langle u, v\rangle$=0

Hence u $\perp$v. Therefore eigenvectors u and v of a symmetric matrix are orthogonal.