bmrm user manual

Julien Prados
University of Geneva

January 15, 2015

1 Introduction

Package *bmrm* implements the "Bundle Methods for Regularized Risk Minimization" proposed by Teo *et al.* (2010). This framework efficiently solves a minimization problem encountred in many recent machine learning algorithm where the goal is to minimze a loss function $l(w, x_i, y_i)$ on the training instances (x_i, y_i), under a regularization term $(\Omega(w))$:

$$\min_w J(w) := \lambda\Omega(w) + R_{emp}(w),$$

$$R_{emp} := \frac{1}{m}\sum l(x_i, y_i, w), \lambda > 0$$

To date, the package implements 10 loss functions providing access to many powerful algorithms with either l1-norm or l2-norm regularization: linear-SVM-classification (with l1 and l2 regularization), multiclass-SVM (with l1 and l2 regularization), epsilon-regression, ordinal-regression, max-margin-fbeta-classification, quantile-regression, etc. Furthermore, flexibility of the framework makes it particulary easy to implement custom loss function for your all your needs.

2 *bmrm* for iris classification

This section shows how to use *bmrm* to train several classification algrithms on *iris* dataset. To simplify the dataset and facilitate plotting, we consider only 2 dimensions (Sepal.Length, Sepal.Width), and limit ourselves to 2 classes (negative class being setosa;positive class being versicolor and virginica).

```
>   require(bmrm)
>   # -- Create a 2D dataset with the first 2 features of iris, and binary labels
>   x <- data.matrix(iris[1:2])
>   y <- c(-1,1,1)[iris$Species]
>   # -- Add a constant dimension to the dataset to learn the intercept
>   x <- cbind(x,1)
```

On this dataset, 3 linear classfiers are learned: linear-SVM with L1-norm regularization, linear-SVM with L2-norm regularization, max-margin-f1-classification with L1-regularization.

```
>   train.prediction.model <- function(...) {
+     m <- bmrm(...)
+     m$f <- x %*% m$w
+     m$y <- sign(m$f)
+     m$contingencyTable <- table(y,m$y)
+     return(m)
+   }
>   # -- train models with maxMarginLoss and fbetaLoss
>   models <- list(
+     svm_L1 = train.prediction.model(hingeLoss(x,y),LAMBDA=0.01,regfun='l1'),
+     svm_L2 = train.prediction.model(hingeLoss(x,y),LAMBDA=0.1,regfun='l2'),
+     f1_L1 = train.prediction.model(fbetaLoss(x,y),LAMBDA=0.01,regfun='l1')
+   )
```

```
1:gap=149.996, loss=150, ub=150, nnz=1
2:gap=149.859, loss=149.88, ub=149.884, nnz=2
3:gap=27.9588, loss=27.9679, ub=27.9927, nnz=2
4:gap=14.7135, loss=14.7226, ub=14.7564, nnz=2
5:gap=14.7082, loss=18.648, ub=14.7564, nnz=2
6:gap=8.38279, loss=8.39307, ub=8.4413, nnz=2
7:gap=8.36929, loss=10.5981, ub=8.4413, nnz=2
8:gap=4.43828, loss=4.48374, ub=4.55576, nnz=2
9:gap=4.43329, loss=12.3863, ub=4.55576, nnz=3
10:gap=4.43007, loss=8.79061, ub=4.55576, nnz=3
11:gap=3.91214, loss=3.92002, ub=4.04571, nnz=3
12:gap=3.90994, loss=4.28462, ub=4.04571, nnz=3
13:gap=2.30553, loss=2.31924, ub=2.45501, nnz=3
14:gap=2.24679, loss=2.87966, ub=2.45501, nnz=3
15:gap=1.69549, loss=1.73893, ub=1.94716, nnz=3
16:gap=1.68618, loss=2.14957, ub=1.94716, nnz=3
17:gap=0.976512, loss=1.02439, ub=1.28537, nnz=3
18:gap=0.743631, loss=0.751592, ub=1.06045, nnz=3
19:gap=0.443874, loss=0.459259, ub=0.776074, nnz=3
20:gap=0.2457, loss=0.3, ub=0.6322, nnz=3
21:gap=0.0979286, loss=0.1, ub=0.4865, nnz=3
22:gap=0, loss=0, ub=0.388571, nnz=3
LowRankQP CONVERGED IN 11 ITERATIONS

    Primal Feasibility    =   3.4492565e-16
    Dual Feasibility      =   0.0000000e+00
    Complementarity Value =   8.8012897e-13
    Duality Gap           =   8.8012888e-13
    Termination Condition =   8.7949960e-13
```

```
1:gap=149.993, loss=150, ub=150, nnz=3
LowRankQP CONVERGED IN 9 ITERATIONS

    Primal Feasibility    =   1.2745523e-14
    Dual Feasibility      =   1.1102230e-16
    Complementarity Value =   2.0698818e-12
    Duality Gap           =   2.0699029e-12
    Termination Condition =   2.0365524e-12
2:gap=149.836, loss=161.096, ub=150, nnz=3
LowRankQP CONVERGED IN 10 ITERATIONS

    Primal Feasibility    =   7.8650930e-14
    Dual Feasibility      =   2.2204460e-16
    Complementarity Value =   2.5296653e-12
    Duality Gap           =   2.5297359e-12
    Termination Condition =   2.4554923e-12
3:gap=27.6785, loss=27.8169, ub=27.9806, nnz=3
LowRankQP CONVERGED IN 16 ITERATIONS

    Primal Feasibility    =   2.0448256e-13
    Dual Feasibility      =   2.2204460e-16
    Complementarity Value =   6.3898796e-13
    Duality Gap           =   6.3896425e-13
    Termination Condition =   6.0779790e-13
4:gap=14.1788, loss=14.3899, ub=14.692, nnz=3
LowRankQP CONVERGED IN 17 ITERATIONS

    Primal Feasibility    =   1.6445867e-13
    Dual Feasibility      =   0.0000000e+00
    Complementarity Value =   5.5796005e-13
    Duality Gap           =   5.5768563e-13
    Termination Condition =   5.2537206e-13
5:gap=14.0717, loss=19.008, ub=14.692, nnz=3
LowRankQP CONVERGED IN 17 ITERATIONS

    Primal Feasibility    =   1.5357620e-13
    Dual Feasibility      =   0.0000000e+00
    Complementarity Value =   1.5092704e-12
    Duality Gap           =   1.5088034e-12
    Termination Condition =   1.3980690e-12
6:gap=7.8874, loss=8.06251, ub=8.68279, nnz=3
LowRankQP CONVERGED IN 18 ITERATIONS

    Primal Feasibility    =   7.4174083e-13
    Dual Feasibility      =   2.2204460e-16
    Complementarity Value =   3.9318662e-12
```

```
    Duality Gap           =   3.9356340e-12
    Termination Condition =   3.4853070e-12
7:gap=7.40153, loss=8.55395, ub=8.68279, nnz=3
LowRankQP CONVERGED IN 18 ITERATIONS

    Primal Feasibility    =   4.1112751e-13
    Dual Feasibility      =   0.0000000e+00
    Complementarity Value =   5.1579073e-11
    Duality Gap           =   5.1579241e-11
    Termination Condition =   3.7137925e-11
8:gap=1.06934, loss=3.6766, ub=4.95786, nnz=3
LowRankQP CONVERGED IN 20 ITERATIONS

    Primal Feasibility    =   5.3278572e-13
    Dual Feasibility      =   0.0000000e+00
    Complementarity Value =   8.3148892e-13
    Duality Gap           =   8.3291707e-13
    Termination Condition =   5.8934418e-13
9:gap=0.849145, loss=3.77062, ub=4.95786, nnz=3
LowRankQP CONVERGED IN 19 ITERATIONS

    Primal Feasibility    =   8.4799407e-13
    Dual Feasibility      =   2.2204460e-16
    Complementarity Value =   6.7549139e-12
    Duality Gap           =   6.7542638e-12
    Termination Condition =   4.7823309e-12
10:gap=0.833129, loss=3.17425, ub=4.95786, nnz=3
LowRankQP CONVERGED IN 19 ITERATIONS

    Primal Feasibility    =   5.5029947e-13
    Dual Feasibility      =   2.2204460e-16
    Complementarity Value =   2.6401054e-11
    Duality Gap           =   2.6402575e-11
    Termination Condition =   1.8456308e-11
11:gap=0.50793, loss=2.47686, ub=4.81255, nnz=3
LowRankQP CONVERGED IN 19 ITERATIONS

    Primal Feasibility    =   4.2736708e-13
    Dual Feasibility      =   4.4408921e-16
    Complementarity Value =   1.1017474e-11
    Duality Gap           =   1.1015161e-11
    Termination Condition =   7.6609094e-12
12:gap=0.431136, loss=2.65696, ub=4.81255, nnz=3
LowRankQP CONVERGED IN 21 ITERATIONS

    Primal Feasibility    =   3.7612416e-13
```

```
    Dual Feasibility      =   1.1102230e-16
    Complementarity Value =   4.4160056e-11
    Duality Gap           =   4.4161785e-11
    Termination Condition =   3.0489601e-11
13:gap=0.328909, loss=2.60099, ub=4.81255, nnz=3
LowRankQP CONVERGED IN 20 ITERATIONS

    Primal Feasibility    =   4.3545565e-13
    Dual Feasibility      =   8.8817842e-16
    Complementarity Value =   1.1841770e-11
    Duality Gap           =   1.1843027e-11
    Termination Condition =   8.1616936e-12
14:gap=0.137949, loss=3.17551, ub=4.64691, nnz=3
LowRankQP CONVERGED IN 20 ITERATIONS

    Primal Feasibility    =   3.1566180e-13
    Dual Feasibility      =   0.0000000e+00
    Complementarity Value =   1.1284388e-12
    Duality Gap           =   1.1272094e-12
    Termination Condition =   7.7404938e-13
15:gap=0.01894, loss=2.98301, ub=4.59732, nnz=3
LowRankQP CONVERGED IN 19 ITERATIONS

    Primal Feasibility    =   2.4928288e-13
    Dual Feasibility      =   3.3306691e-16
    Complementarity Value =   6.1874133e-12
    Duality Gap           =   6.1876060e-12
    Termination Condition =   4.2418067e-12
16:gap=0.0105815, loss=3.07374, ub=4.59732, nnz=3
LowRankQP CONVERGED IN 18 ITERATIONS

    Primal Feasibility    =   2.7018400e-13
    Dual Feasibility      =   3.3306691e-15
    Complementarity Value =   4.1465469e-11
    Duality Gap           =   4.1466997e-11
    Termination Condition =   2.8426823e-11
17:gap=-3.8785e-11, loss=3.03087, ub=4.58674, nnz=3
1:gap=0.999987, loss=1, ub=1, nnz=1
2:gap=0.865809, loss=0.865869, ub=0.865882, nnz=2
3:gap=0.203569, loss=0.203581, ub=0.203654, nnz=2
4:gap=0.0731116, loss=0.0731403, ub=0.0732258, nnz=2
5:gap=0.0501698, loss=0.0501917, ub=0.0503059, nnz=2
6:gap=0.0341375, loss=0.0341776, ub=0.0343138, nnz=2
7:gap=0.0282367, loss=0.0282915, ub=0.0284679, nnz=2
8:gap=0.0211995, loss=0.0212955, ub=0.0215267, nnz=3
9:gap=0.0211765, loss=0.0545747, ub=0.0215267, nnz=3
```

```
10:gap=0.0211292, loss=0.0236855, ub=0.0215267, nnz=3
11:gap=0.016833, loss=0.0168427, ub=0.0172402, nnz=3
12:gap=0.0168063, loss=0.0280392, ub=0.0172402, nnz=3
13:gap=0.00844274, loss=0.00846368, ub=0.00889756, nnz=3
```

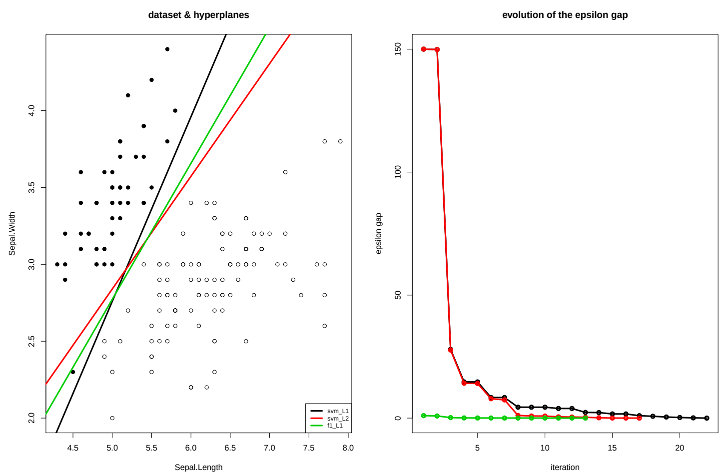

Figure 1: Left panel: Comparison of the decision surface of the 3 linear models trained on iris dataset. Right panel: Convergence curve of the optimization process

References

Teo CH, Vishwanathan S, Smola A, Le QV (2010). "Bundle Methods for Regularized Risk Minimization." *Journal of Machine Learning Research*, **11**, 311–365.