bmrm user manual

Julien Prados
Genetic Psychiatric Unit, University of Geneva

September 4, 2013

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(x,y,lossfun=hingeLoss,...) {
+   m <- bmrm(x,y,lossfun=lossfun,...)
+   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(x,y,lossfun=hingeLoss,LAMBDA=0.01,regfun='l1'),
+   svm_L2 = train.prediction.model(x,y,lossfun=hingeLoss,LAMBDA=0.1,regfun='l2'),
+   f1_L1 = train.prediction.model(x,y,lossfun=fbetaLoss,LAMBDA=0.01,regfun='l1')
+ )
```

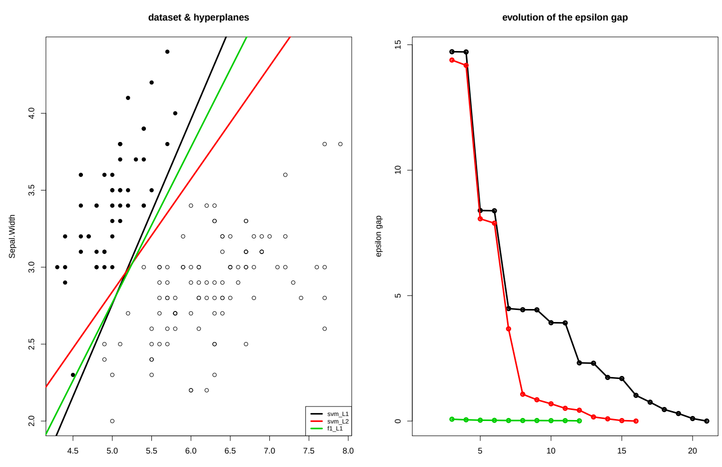

Figure 1: Left panel: Comparison of the decision surface of the 3 linear models trained on iris dataset. Right panel: Convergence curve of the optimization process

References

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