

# CSI 5325 Assignment 4

Greg Hamerly

Assigned: 3/13/2018; Due: 3/29/2018

## Instructions

The instructions for this assignment are the same as for all assignments in this course; for details refer to Assignment 1. As a refresher: use  $\text{\LaTeX}$ , make your document beautiful (proofread it!), use well-labeled figures to illustrate things, turn in a hardcopy and an email copy, and keep attachments small.

## Problems

Do the following exercises. Do not just give your answers; show your work (in  $\text{\LaTeX}$ ) and explain/analyze your results.

1. Implement LASSO and Ridge regression. Recall from class that these minimize the Lagrangian functions, or minimize the constrained in-sample error:

Method	Lagrangian formulation	Constraint formulation
Ridge	$\min_w E_{\text{in}}(w) + \lambda \sum_{i=1}^d w_i^2$	$\min_w E_{\text{in}}(w)$ subject to $\sum_{i=1}^d w_i^2 \leq C$
LASSO	$\min_w E_{\text{in}}(w) + \lambda \sum_{i=1}^d  w_i $	$\min_w E_{\text{in}}(w)$ subject to $\sum_{i=1}^d  w_i  \leq C$

- (a) Show the mathematics to minimize the Ridge regression (Lagrangian formulation only) and the LASSO regression (Constraint formulation only). For Ridge, you can show the gradient descent updates, or you can show how to compute it in one shot. For LASSO, show the gradient descent update.

Recall in class we discussed a method for handling when the budget constraint is violated. We project the LASSO solution back onto the surface of the budget by doing the following:

- Let  $w$  be the current weight vector which violates the budget.
- Define the vector  $z$  as pointing in the same orthant as  $w$ , specifically  $z_i = \text{sign}(w_i)$ . Let  $s = \sum_i |z_i|$ . Then update  $z = z \times C/s$ .
- Let  $o = ((w^T z)/(z^T z) - 1) \times z$  be the offset between  $w$  and the projection surface.
- Replace  $w$  with  $w - o$ .
- Repeat as necessary until the budget constraints hold.

As an alternative to the above method, you could follow Rob Tibshirani's advice on using least-angle regression to solve LASSO, which has a few details here: <http://statweb.stanford.edu/~tibs/lasso/simple.html>. You could even compare the two methods.

- (b) Use both LASSO and Ridge regression to fit (regularized) polynomials of order 10 (e.g.  $h(x) = w_0 + w_1x + w_2x^2 + \dots + w_{10}x^{10}$ ). You can generate data with the following function, or make your own interesting generator.

```
function [x, y] = genData(n)
    x = sort(randn(n, 1)) * 2;
    f = (cos(x) + 2) ./ (cos(1.4 .* x) + 2);
    noise = randn(n, 1) * std(f) * 0.5;
    y = f + noise;
```

Generate both training and test datasets. Use gradient descent to train LASSO weights and gradient descent or the pseudoinverse to train Ridge regression weights. Since regularization is most effective when there is noise but few training points, use small values for  $n$  to see the difference that regularization makes.

One important aspect to getting gradient descent to behave well is ensuring that all features are in the same range. Thus, you should probably standardize (i.e. z-score) each of the non-constant inputs ( $x$ ,  $x^2$ , etc.), as well as the output  $y$ .

- (c) Plot the in-sample and out-of-sample error as a function of  $\lambda$  (for Ridge) or  $C$  (for LASSO). Compare the sparseness of the weights obtained using both methods.
2. Using the same dataset you just tried, implement both non-parametric and parametric RBFs. For non-parametric, refer to Equation (6.1) in LFD. For parametric, refer to Equation (6.4). In the parametric case, vary  $k$ , and place the centers at regular intervals in the input space. For both parametric and non-parametric models, vary  $r$ . Show how both  $k$  and  $r$  affect the shape of the fit and  $E_{\text{test}}$ . How do the book's suggestions (rules of thumb) for  $k$  and  $r$  work for you?
3. LFD Problem 6.14. Use the data provided at <http://amlbook.com/support.html> (the preprocessed symmetry and intensity dataset).