CS492(E) Homework1
Optimization problem and Linear Regression

(Due Date: 2017-11-07 11:59 pm)

Dataset. In this programming homework, we will use a LIBSVM dataset which are pre-processed data originally from UCI data repository.

- Housing dataset (We will use housing_scale dataset). Predict housing values in suburbs of Boston.
  https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression/housing_scale.

$$y = \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x} + \beta_0 + \epsilon,\ \ \epsilon: noise - (1) \text{ where } \boldsymbol{x} = (x_1, \dots, x_p), \boldsymbol{\beta} = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$$

Problem 1. Randomly split the dataset into two groups: training (around 80%) and testing (around 20%). Learn the linear regression model on the training data, using the analytic solution. Compute the prediction error on the test data: $\frac{1}{\text{\# of test points}} \Sigma_i |\widehat{y_i} - y_i^*|$ where $\widehat{y_i}$ and y_i^* are the prediction and the true value respectively for data point i. Repeat this process 10 times and report all individual prediction errors of 10 trials and the average of them. How are the results changed when we assume bias term β_0 of linear regression model is zero?

- (Answer to the following question) For fixed weight $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, what is the value of $\beta_0 = \beta_0^*$ that minimizes the objective function of linear regression model (eq. 1)? Explain what the role of the bias parameter β_0 is.

Problem 2. Do the same work as in the problem #1 but now using a gradient descent (10 randomly generated datasets in Problem 1 should be maintained; we will use the datasets generated in Problem1.) Here we are not using (exact or backtracking) line searches. You need to try several selections for the fixed step size.

(a) Compare prediction errors with those from Problem 1

(b) Additionally, draw plots showing objective function values vs. iterations of gradient descent. In all plots, optimal objective function value by analytic solution should be presented (it would be a horizontal line, as shown in Figure 1). Report for different step sizes (for too large, proper and too small step sizes).

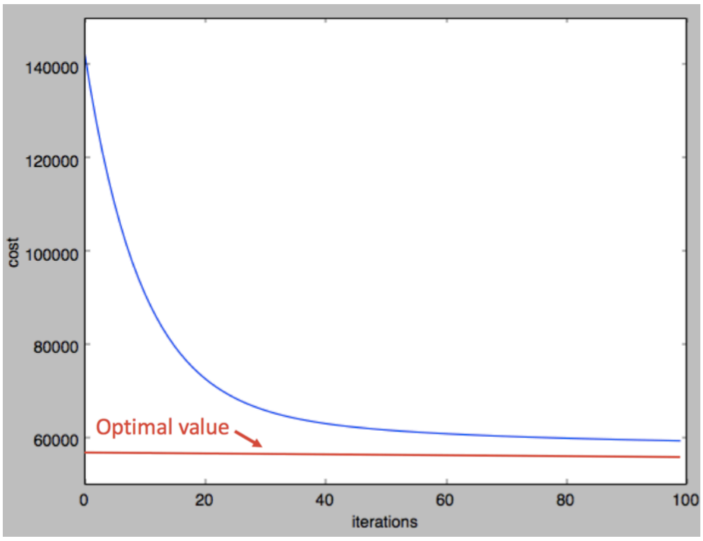

Figure 1 : Example plot.

Problem 3. Do the same work as in the problem #2 but now using a coordinate descent (10 randomly generated datasets in Problem 1 should be maintained; we will use the datasets generated in Problem1.)

(a) Compare prediction errors with those from previous problems

(b) Additionally, draw plots showing objective function values vs. iterations of a coordinate descent. In all plots, optimal objective function value by analytic solution should be presented (it would be a horizontal line, as shown in Figure 1).

Problem 4 (weighted error) Consider a dataset in which each data point y_n is associated with a weighting factor $r_n > 0$, so that the sum-of-squares error function becomes

$$E_D(\boldsymbol{\beta}) = \frac{1}{2}\Sigma_{n=1}^{N} r_n \{y_n - \boldsymbol{\beta}^T \boldsymbol{x}_n\}^2$$

Find an expression for the solution $\boldsymbol{\beta}^*$ that minimizes this error function. Give two alternative interpretations of the weighted sum-of-squares error function in terms of (i) data dependent noise variance and (ii) replicated data points.

Problem 5 (convex function) Is the function $\text{logSumExp}(x_1, x_2, \ldots, x_n) = \log(\exp(x_1) + \exp(x_2) + \cdots + \exp(x_n))$ *convex*? Justify your answer.

Problem 6 (dual problem) Consider the following linear programming

$$p^* \coloneqq \min \boldsymbol{c}^T \boldsymbol{x} \ \text{ s.t } \ Ax \le b$$

Determine the dual problem form by using Lagrangian multipliers.