EECS289A Introduction to Machine Learning, Project T Final - Quiz Questions

Han Liu, Peng Tan, Dilu Xu, Jinyan Zhao
Department of Civil Engineering
Universitu of California, Berkeley
Berkeley, CA 94704
{han_liu, tanpeng, diluxu, jinyan_zhao}@berkeley.edu

November 26, 2020

1 Elastic Net (Han)

In class we discussed two types of regularization, l_1 and l_2. Both are useful, and sometimes it is helpful combine them, giving the objective function below (**in this problem we have excluded w_0 for simplicity**):

$$F(\boldsymbol{w}) = \frac{1}{2}\sum_{j=1}^{n}(y^{(j)} - \sum_{i=1}^{d}\boldsymbol{w}_i\boldsymbol{x}_i^{(j)})^2 + \alpha\sum_{i=1}^{d}|\boldsymbol{w}_i| + \frac{\lambda}{2} + \sum_{i=1}^{d}\boldsymbol{w}_i^2 \tag{1}$$

Here, $(\boldsymbol{x}^{(j)}, y^{(j)})$ is j-th example in the training data, w is a d dimensional weight vector, λ is a regularization parameter for the l_2 norm of w, and α is a regularization parameter for the l_1 norm of w. This approach is called the Elastic Net, and you can see that it is a generalization of Ridge and Lasso regression: It reverts to Lasso when $\lambda = 0$, and it reverts to Ridge when $\alpha = 0$. In this question, we are going to derive the coordinate descent (CD) update rule for this objective.

Let g, h, c be real constants, and consider the function of x

$$f_1(x) = c + gx + \frac{1}{2}hx^2 \tag{2}$$

1. [4 points] What is the x^* that minimizes $f_1(x)$? (i.e. calculate $x^* = \arg\min f_1(x)$)
Answer:
Take the gradient of $f_1(x)$ and set it to 0.

$$g + hx = 0 \tag{3}$$

$$x^* = -\frac{g}{h} \tag{4}$$

Let α be an additional real constant, and consider another function of x

$$f_2(x) = c + gx + \frac{1}{2}hx^2 + \alpha|x|(h > 0, \alpha > 0) \tag{5}$$

This is a piecewise function, composed of two quadratic functions:

$$f_2^-(x) = c + gx + \frac{1}{2}hx^2 - \alpha x \tag{6}$$

and

$$f_2^+(x) = c + gx + \frac{1}{2}hx^2 + \alpha x \tag{7}$$

Let $\tilde{x}^- = \arg\min_{x \in \mathbb{R}} f_2^-(x)$ and $\tilde{x}^+ = \arg\min_{x \in \mathbb{R}} f_2^+(x)$.
(**Note:** The argmin is taken over $(-\infty, +\infty)$ here.