6.867: Homework 2

1. Logistic regression

In this section, we explore logistic regression with L1 and L2 regularization. We use gradient descent to compare the resulting weight vectors under different regularizers and regularization parameters, and we evaluate the effect of these choices in the context of multiple data sets.

1.1. L2 regularization

We first consider L2 regularization, in which the objective function to minimize is

$$E_{LR}(w, w_0) = \text{NLL}(w, w_0) + \lambda |w|^2$$

where

$$\text{NLL}(w, w_0) = \sum_i \log(1 + \exp(-y^{(i)}(wx^{(i)} + w_0)))$$

and in the case of L2 regularization,

$$|w| = |w|_2 = \sqrt{w_1^2 + ... + w_n^2}$$

Gradient descent was run with this objective function on the training dataset `data1_train.csv` with $\lambda = 0$. Interestingly, as the number of gradient descent iterations, controlled by the step size and the convergence criterion, increased, the weight vector _____

1.2. L1 regularization

In the case of L1 regularization, we get that in the above equation,

$$|w| = |w|_1 = \sum_{i=1}^{n} |w_i|$$

We can evaluate the different regularization techniques under different values of λ in the context of the weight vectors, the decision boundary, and the classification error rate in each of the training data sets.

Dataset	Best regularizer	Best λ	Test performance
1	5	6	1
2	5	6	1
3	5	6	1
4	5	6	1

Table 1. Optimal regularizer and λ for datasets

1.2.1. WEIGHT VECTOR

1.2.2. DECISION BOUNDARY

1.2.3. CLASSIFICATION ERROR RATE

1.3. Optimization

By using the training and validation data sets, we can identify the best regularizer and value for λ for each of the four data sets. These results are presented in Table 1.

2. Support Vector Machine

3. Pegasos

4. Handwritten digit recognition