
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,

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

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 -----