Introduction to Statistical Machine Learning
Problem set 4

Alan Joseph Bekker

Multi-class Case

Given $x_1, ..., x_n \in R^d$ and the multiclass labels $y_1, ..., y_n \in 1, ...k$ the cost function we optimize is:

$$l(w) = \sum_{t=1}^{n} \log p(y_t|x_t) = \sum_{t=1}^{n} \log \sum_{i=1}^{k} 1_{\{y_t=i\}} p(y_t = i|x_t; w)$$

logistic regression model: $\hat{y}_i = p(y = i|x; w) = \frac{\exp(w_i x)}{\sum_{j=1}^{k} \exp(w_j x)}$

Note in Multi-class logit we have k different vectors w_i, in contrast to the binary case we have only one.

Gradient ascent algorithm:

For $i = 1, ...k$

$$w_i \leftarrow w_i + \epsilon \frac{1}{N} \sum_t (1_{\{y_t=i\}} - \hat{y}_{ti}) x_t$$

1.
   - If you have already download the MNIST data, neglect this: Download mnistSmall.mat from: `https://github.com/dustinstansbury/medal/blob/master/data/mnistSmall.mat`
   - Visualize the data using the "imshow(reshape(testData(1,:),[28,28]))" command and verify you understand how is the data looks like.
   - Train a Multi-Class Logistic Regression algorithm using the Gradient Ascent method described above. Train the machine to distinguish between all the digits.
   - Print the Cost Function l_w at each iteration of the optimization procedure and verify it increases.
   - What is the Sucsses Rate you achieved?