Introduction to Machine learning
Problem set 3

Alan Joseph Bekker

The logistic regression model is $p(y=1|x) = 1 - p(y=0|x) = g(wx)$ where $g(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function. Given $x_1, ..., x_n \in R^d$ and binary labels $y_1, ..., y_n$, the cost function we optimize is:

$$l(w) = \frac{1}{N}\sum_{t=1}^{N} \log p(y_t|x_t)$$

Gradient ascent: $w_i \leftarrow w_i + \tau \frac{\partial l(w)}{\partial w_i}, \qquad i = 0, ..., d$

1.
- Show that $g'(z) = g(z)(1 - g(z))$.
- Prove that

$$\frac{\partial l(w)}{\partial w} = \frac{1}{N}\sum_{t=1}^{N}(y_t - g(wx_t))\vec{x_t}$$

  Hint: observe that $p(y|x) = g(wx)^y(1 - g(wx))^{(1-y)}$
- Download the attached data.
- Visualize some of the feature vectors, using reshape((28,28)) and be sure you understand the data.
- Train a binary Logistic Regression algorithm using the Gradient Ascent method described above. Train the machine to distinguish between the digits 1 & 2.
- Print the Cost Function l_w at each iteration of the optimization procedure and verify it increases.
- What is the Sucsses Rate you achieved?