Machine Learning, 2024 Spring
Assignment 3

Notice

Plagiarizer will get 0 points.
LaTeX is highly recommended. Otherwise you should write as legibly as possible.

Problem 1 For logistic regression, show that

$$\nabla E_{\text{in}}(\mathbf{w}) = -\frac{1}{N}\sum_{n=1}^{N}\frac{y_n\mathbf{x}_n}{1+e^{y_n\mathbf{w}^{\mathrm{T}}\mathbf{x}_n}} = \frac{1}{N}\sum_{n=1}^{N}-y_n\mathbf{x}_n\theta\left(-y_n\mathbf{w}^{\mathrm{T}}\mathbf{x}_n\right) \tag{1}$$

Argue that a 'misclassified' example contributes more to the gradient than a correctly classified one.
Solution

Problem 2 For linear regression, the out-of-sample error is

$$E_{\text{out}}(h) = \mathbb{E}[(h(x) - y)^2]. \tag{2}$$

Show that among all hypotheses, the one that minimizes E_{out} is given by

$$h^*(x) = \mathbb{E}[y|x]. \tag{3}$$

The function h^* can be treated as a deterministic target function, in which case we can write $y = h^*(x) + \epsilon(x)$ where $\epsilon(x)$ is an (input dependent) noise variable. Show that $\epsilon(x)$ has expected value zero.

Solution

Lamma: The law of iterated Expectation

$$\mathbb{E}(Y) = \mathbb{E}_X[\mathbb{E}(Y|X)]$$

$$\begin{aligned}
LHS &= \mathbb{E}(Y) = \int_{-\infty}^{+\infty} y f_y(y) dy \\
RHS &= \mathbb{E}_X \left[\int_{-\infty}^{+\infty} y \frac{f(x,y)}{f_x(x)} dy \right] \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} y \frac{f(x,y)}{f_x(x)} dy \right] f_x(x) dx \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x,y) dx dy \\
&= \int_{-\infty}^{+\infty} y \left[\int_{-\infty}^{+\infty} f(x,y) dx \right] dy \\
&= \int_{-\infty}^{+\infty} y f_y(y) dy \\
&= LHS
\end{aligned}$$

Problem 3 According to the iterative format provided as follow:

- Use the SUV dataset to implement (using Python or MATLAB) the Gradient Descent method to find the optimal model for logistic regression.
- What is your test error? What are the sizes of the training set and test set, respectively?
- What is your learning rate? How was it chosen? How many steps were iterated in total?
- Present the results of the last 10 steps produced by your algorithm, including the loss, learning rate, the L2 norm of the gradient, and the number of function evaluations and gradient evaluations.

Fixed learning rate gradient descent:

1: Initialize the weights at time step $t = 0$ to $\mathbf{w}(0)$.
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: Compute the gradient $\mathbf{g}_t = \nabla E_{\text{in}}(\mathbf{w}(t))$.
4: Set the direction to move, $\mathbf{v}_t = -\mathbf{g}_t$.
5: Update the weights: $\mathbf{w}(t+1) = \mathbf{w}(t) + \eta \mathbf{v}_t$.
6: Iterate to the next step until it is time to stop.
7: Return the final weights.

Dataset reference and description
Dataset and download

Solution