CS589: Machine Learning - Fall 2017

Quiz 3

Oct. 23, 2017 Name: Jie Wang

Instructions: There are 14 total questions each with 7 points, with 2 points free. Only the final answer for each question will be graded, with no partial credit. If you do intermediate calculations, please draw a box around your final answer. Each minute late the quiz is turned in will result in a 20 point penalty.

1. A Scenario For Classification

Setup

Take a classification dataset $(x_1, y_1), ..., (x_N, y_N)$ where $x \in \mathbb{R}^p$ and a corresponding validation set $(x_1, y_1), ..., (x_M, y_M)$. The output is binary, with $y_i \in \{-1, +1\}$, and you have $N = 100$ and $p = 5$. You train a linear model using hinge loss with ridge regularizarion and a ridge penalty of $\lambda = 0.1$. After training, you observe a mean training hinge loss of 0.3 and a mean validation hinge loss of 0.4. (Regularization penalties are used while training, but not included in the losses shown here.)

Questions

1. Is it possible to observe a mean train 0-1 loss of 0.07? **Possible / Not possible.**
2. Is it possible to observe a mean validation 0-1 loss of 0.07? **Possible / Not possible.**

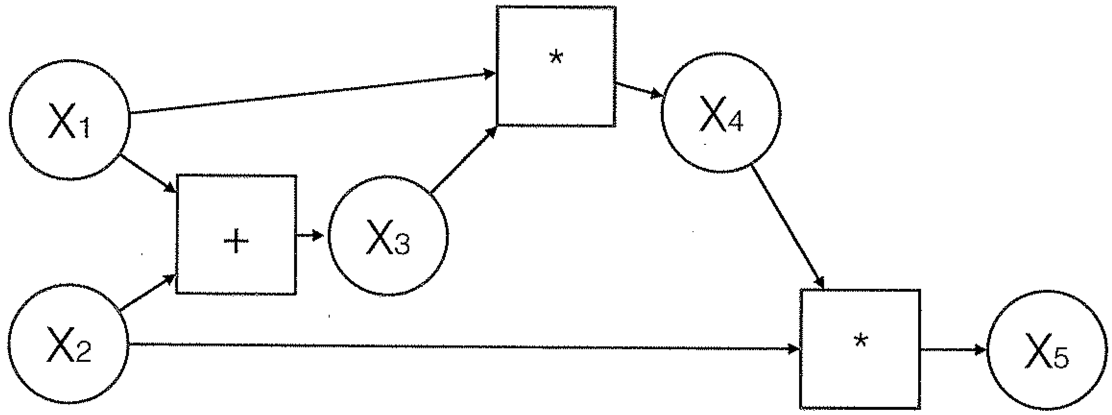
Change: Now, you re-train the model using hinge loss, but no regularization, i.e. $\lambda = 0$.

Questions

3. Is it possible to now observe a mean train hinge loss of .07? **Possible / Not possible.**
4. Is it possible to now observe a mean validation hinge loss of .07? **Possible / Not possible.**
5. Is it possible to now observe a mean train 0-1 loss of .07? **Possible / Not possible.**
6. Is it possible to now observe a mean validation 0-1 loss of .07? **Possible / Not possible.**

2. Autodiff

In this problem, you will perform automatic differentiation on the function $f(x_1, x_2)$ defined by the following expression graph. Thus, f takes x_1 and x_2 as inputs, and returns x_5. (Here, " + " means addition and " $*$ " means multiplication.)



Given an input of $x_1 = 1.0$ and $x_2 = 1.0$ what does the network evaluate to during forward propagation?

$x_3 = X_1 + X_2 = 1.0 + 1.0 = 2.0$

$x_4 = X_1 X_3 = 1.0 \times 2.0 = 2.0$

$x_5 = X_2 \cdot X_4 = 1.0 \times 2.0 = 2.0$

Now, what does the network evaluate to during back propagation?

$\frac{df}{dx_5} = 1$

$\frac{df}{dx_4} = \sum_{j:4 \in Pa(j)} \frac{df}{dx_j} \frac{dg_j}{dx_4} = \frac{df}{dx_5} \frac{dx_5}{dx_4} = 1 \cdot X_2 = 1$

$\frac{df}{dx_3} = \frac{df}{dx_4} \frac{dg_4}{dx_3} = 1 \cdot X_1 = 1$

$\frac{df}{dx_2} = \frac{df}{dx_3} \frac{dg_2}{dx_2} + \frac{df}{dx_5} \frac{dg_5}{dx_2} = 1 \cdot 1 + 1 \cdot X_4 = 1 + 1 \cdot 2 = 3$

$\frac{df}{dx_1} = \frac{df}{dx_3} \frac{dg_2}{dx_1} + \frac{df}{dx_4} \frac{dx_4}{dx_1} = 1 \cdot 1 + 1 \cdot X_3 = 1 + 2 = 3$